

**Project Report  
HTAP-9**

# **A Taxonomy of Spectral Unmixing Algorithms and Performance Comparisons**

**N. Keshava**

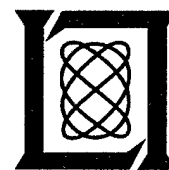
**15 January 2002**

---

**Lincoln Laboratory**

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

*LEXINGTON, MASSACHUSETTS*



---

Prepared for the Department of the Under Secretary of Defense,  
S&T, under Air Force Contract F19628-00-C-0002.

Approved for public release; distribution is unlimited

**20020129 051**

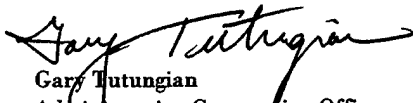
This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Under Secretary of Defense, S&T, under Air Force Contract F19628-00-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

  
Gary Tutungian  
Administrative Contracting Officer  
Plans and Programs Directorate  
Contracted Support Management

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document  
when it is no longer needed.

Massachusetts Institute of Technology  
Lincoln Laboratory

**A Taxonomy of Spectral Unmixing Algorithms  
and Performance Comparisons**

*N. Keshava*  
*Group 97*

Project Report HTAP-9

15 January 2002

Approved for public release; distribution is unlimited

## ABSTRACT

In this report, algorithms for spectral unmixing are organized into taxonomies and their performance is then compared. Our motivation is to collectively organize and relate algorithms in order to assess the current state-of-the-art in the field and to facilitate objective comparisons between methods. The hyperspectral sensing community is populated by investigators with disparate scientific backgrounds and efforts in spectral unmixing developed within disparate communities have inevitably led to duplication. This report is intended to remove ambiguity and redundancy by using a standard vocabulary, and clearly summarize what has and has not been done. As will be evident, the framework for the taxonomies derives its organization from the fundamental, philosophical assumptions imposed on the problem, rather than the common calculations they perform, or the similar outputs they might yield. The taxonomies are supplemented by a comparison of unmixing performance using techniques that typify the approaches of wide classes of algorithms.

## ACKNOWLEDGMENTS

The author would like to thank CAPT Frank Garcia at DUSD (S&T), program manager for the Lincoln Laboratory Hyperspectral Technology Assessment Program (HTAP) for supporting this effort. The author would like to thank John Kerekes, Dimitris Manolakis, and Gary Shaw for their contributions to this work. The author would like to thank Craig Richard for performing the atmospheric compensation, David Marden for advice on computing issues, and Peter Boettcher for implementing the geometric endmember determination algorithm. In addition, the author extends his gratitude to Andy McKellips for providing a LaTeX template for this report.

## TABLE OF CONTENTS

	Page
Abstract	iii
Acknowledgments	v
List of Illustrations	ix
 1. INTRODUCTION	 1
1.1 Motivation	1
1.2 Document Outline	2
 2. PHYSICAL MODELS FOR SPECTRAL MIXING	 3
2.1 Linear Mixing	3
2.2 Non-Linear Mixing	3
 3. ALGORITHM TAXONOMIES: LINEAR UNMIXING	 7
3.1 Taxonomy Structure	7
3.2 Dimension Reduction Taxonomy	8
3.3 Endmember Determination Taxonomy	12
3.4 Inversion Taxonomy	20
 4. ALGORITHM TAXONOMY: NON-LINEAR UNMIXING	 31
 5. PERFORMANCE COMPARISONS	 33
5.1 Data	33
5.2 Selected Algorithms	35
5.3 Approach	35
5.4 Scene Statistics	35
5.5 Endmember Determination	35
5.6 Inversion	39
 6. SUMMARY	 43
 Glossary	 45
References	47

## LIST OF ILLUSTRATIONS

Figure No.		Page
1	Alternate mixing paradigms: (a) linear mixing from a checker-board mixture having a single reflection, (b) non-linear mixing from an intimate mixture having multiple reflections.	4
2	Conceptual structure for taxonomies.	8
3	Taxonomy of dimension reduction algorithms.	9
4	Taxonomy of endmember determination algorithms.	13
5	Taxonomy of inversion algorithms.	21
6	Confidence measurement for 50%-50% mixture (triangles) and 30%-70% (circles) mixture.	23
7	Alpine Radiance Scene I. The data was collected around noon on September 22, 1997.	34
8	(a) Covariance of scene; (b) Cumulative, normalized eigenvalues; (c) First five principal components	36
9	Endmembers for geometric technique using (a) 3-cornered simplex; (b) 4-cornered simplex; (c) 5-cornered simplex; (d) 6-cornered simplex.	38
10	(a) Endmembers for fuzzy clustering algorithm with 3 classes; (b) abundance plane for Endmember 1; (c) abundance plane for Endmember 2; (d) abundance plane for Endmember 3.	40
11	Abundance planes (a) Endmember 1; (b) Endmember 2; (c) Endmember 3; (d) Endmember 4; (e) Endmember 5; (f) Sum of abundances.	41

# 1. INTRODUCTION

Hyperspectral unmixing is the procedure by which pixel spectra in a scene are decomposed into constituent substances and the fractions in which they appear. The collection of constituent members is comprised of materials found in the scene that are deemed pure substances. In the strict sense, constituent members (or endmembers) can be unique elements, (e.g., iron, copper), but in the practical sense of hyperspectral imaging, the endmembers more likely represent disparate macroscopic entities (e.g., sand, trees, vehicle paint, asphalt, concrete).

Each pixel in a hyperspectral image consists of observations of the surface radiance in adjacent electromagnetic channels. After calibration and atmospheric compensation, the observations convey the total spectral response emanating from materials within each pixel. Unmixing algorithms attempt to extrapolate the components that comprise each mixed pixel, as well as the relative proportions in which they appear. The approaches vary in their methods, and through their algorithmic formulations, implicitly incorporate philosophical assumptions regarding the physical mechanisms and mathematical structure by which the reflectance properties from disparate substances combine to yield mixed pixel spectra. Hence, when a mixed pixel is processed by different unmixing algorithms, the disparity of assumptions is the cause of the range of output values.

Understanding the philosophical differences between algorithms is important when comparing the performance of different algorithms. Thus, an algorithm taxonomy provides the first step toward benchmarking algorithm performance and establishing the conditions under which some algorithms fail or succeed.

## 1.1 MOTIVATION

The motivation for writing this report is to provide a clear technical understanding of the universe of spectral unmixing algorithms and to provide a comparison of the performance of algorithms with actual hyperspectral data. Hyperspectral sensing and processing, unlike most other sensing paradigms, resides at the confluence of several disciplines, including, but not limited to, physics, mathematics, engineering, and statistics. It is not coincidental that algorithms for unmixing hyperspectral data emphasize the strengths and biases of the developers. But, the lack of a common vocabulary has resulted in many duplicate efforts. In this sense, the taxonomies strive to distill the universe of known algorithms to a minimal set.

The taxonomies hierarchically organize the algorithms into general categories based on their operating assumptions. Moreover, because the taxonomies reveal that large numbers of algorithms within the same class are often derived from a single underlying formulation, the comparison of their performance can be limited to a single representative, or "best-of-class", algorithm. Moreover, we hope the taxonomies provide a framework for future algorithm development, clearly indicating what has and has not been done. As will become evident in this document, algorithms have been organized at the highest levels by the general assumptions they implicitly impose on the problem. With this in mind, new algorithms should fit into this structure.



## 1.2 DOCUMENT OUTLINE

The principle contributions of this report are three taxonomies (Figures 3, 4, and 5) for spectral unmixing algorithms and comparisons of performance for several algorithms. Section 2 describes two fundamental models for spectral mixing, linear mixing and non-linear mixing. Section 3 introduces the taxonomies for three procedures crucial to linear unmixing, dimensionality reduction, endmember determination, and inversion, and provides a technical discussion of algorithms. Section 4 discusses non-linear unmixing algorithms. In Section 5, the results of performance comparisons between different algorithms are presented. Finally, Section 6 summarizes the findings.

## 2. PHYSICAL MODELS FOR SPECTRAL MIXING

Any approach for effectively unmixing hyperspectral data must begin with a model describing how constituent substances in a scene combine to form the composite spectrum appearing in a received pixel. Mixing models attempt to represent the underlying physics that are the foundation of hyperspectral phenomenology; unmixing algorithms use these models to perform the inverse operation, attempting to recover the constituent spectra and their associated fractional abundances from the mixed pixel spectrum. We present two mixing models that each offer a different interpretation of how incident radiation interacts with an incident surface. Discussions focusing on the mixing models from both an engineering [21] and physics [8,13,14] perspective are available in the literature.

### 2.1 LINEAR MIXING

In Figure 1(a), the reflecting surface is portrayed as a *checkerboard* mixture, and the incident radiation only bounces once on its surface. If the total surface area is conceived to be divided proportionally according to the fractional abundances of the constituent substances, then the reflected radiation will convey with the same proportions the characteristics of the associated media. In this sense, there exists a *linear* relationship between the fractional abundance of the substances comprising the area being imaged, and the spectra in the reflected radiation. Hence, this is called the linear mixing model (LMM), and is expressed as

$$\mathbf{x} = \sum_{k=1}^P a_k \mathbf{s}_k + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w} \quad (1)$$

where  $\mathbf{x}$  is the  $M \times 1$  received pixel spectrum vector,  $\mathbf{s}_k$  is the  $k$ -th  $M \times 1$  column of  $\mathbf{S}$ ,  $\mathbf{a}$  is the  $P \times 1$  fractional abundance vector,  $\mathbf{w}$  is the  $M \times 1$  additive observation noise vector,  $\mathbf{S}$  is the  $M \times P$  matrix whose columns are  $\mathbf{s}_k$ ,  $M$  is the number of spectral bands, and  $P$  is the number of endmembers. Frequently, block notation will be employed using upper-case variables to extend (1) when multiple pixels are considered so that  $\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{W}$ . Then for  $N$  pixels,  $\mathbf{X}$  is a  $M \times N$  matrix whose  $N$  columns are the  $M \times 1$  pixel spectra. Likewise,  $\mathbf{A}$  is a  $P \times N$  matrix of abundances, and  $\mathbf{W}$  is a  $M \times N$  matrix of additive noise values.

### 2.2 NON-LINEAR MIXING

Figure 1(b) depicts a more complicated scenario. The arrangement of the constituent substances is not as orderly as in Figure 1(a) because the substances comprising the medium are not organized proportionally on the surface. This *intimate* mixture results when each component is randomly distributed in a homogeneous way. As a result, the incident radiation can experience reflections with multiple substances, and the aggregate spectrum of reflected radiation may no longer uphold the linear proportions (either in mass fraction or in volume) of the constituent substance spectra. Because (1) is inappropriate to describe this interaction, this scenario is referred to as non-linear mixing. We briefly discuss two non-linear models for spectral mixing.

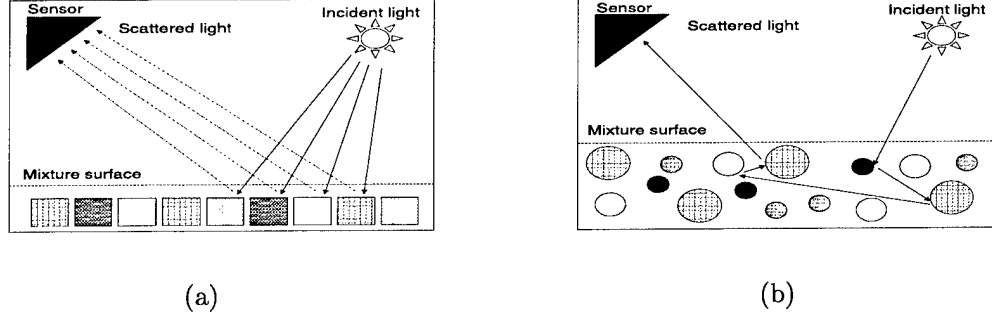


Figure 1. Alternate mixing paradigms: (a) linear mixing from a checkerboard mixture having a single reflection, (b) non-linear mixing from an intimate mixture having multiple reflections.

### 2.2.1 Two Stream Method

Many investigators have demonstrated that vegetation and soil surfaces are not well-described by linear mixing models, and non-linear models are required to describe these kinds of mixtures, which are treated as *particulate media*. One approach develops a model for scattering from particulate surfaces which, in contrast to the linear mixing model, incorporates multiple scattering into the expression for the bidirectional reflectance function (BRDF) for a surface consisting of particles of arbitrary shape in close proximity to one another [13]. The collection of effects induced by multiple scattering in a homogeneous medium can be organized under a single parameter known as single-scattering albedo,  $w$ , which is the probability that a photon survives interactions such as Fresnel reflection, absorption, scattering, and diffraction due to the presence of a single grain of a substance. In the presence of multiple substances, the average single-scattering albedo,  $\bar{w}$ , of a homogeneously mixed, multi-component surface can be approximated at each wavelength by an expression depending on the fractional abundances in the mixture as [8]

$$\bar{w} = \frac{\sum_{k=1}^P \frac{a_k}{\rho_k D_k} w_k}{\sum_{k=1}^P \frac{a_k}{\rho_k D_k}} \quad (2)$$

where  $\rho_k$  is the density of the  $k$ -th substance,  $D_k$  is the  $k$ -th mean grain diameter, and  $w_k$  is the single-scattering albedo for the  $k$ -th substance.

### 2.2.2 Radiosity

An altogether different approach to modeling non-linear mixing employs a more macroscopic viewpoint than the one for particulate media in 2.2.1 [4]. In contrast to the two-stream model in Section 2.2.1 which requires scattering parameters of particles to ascertain the single-scattering albedo and the fractional abundances of particulate mixtures, the method of radiosity balances equations representing the transmitted and reflected radiation between specific geometries of soil

and vegetation; the equations can be solved for the proportion of a specific substance in the model that reconciles theory with the collected data. The solution to the radiosity equation has been formulated for simple canopy geometries such as single-layers of vegetation above soil where multiple reflections and transmissions occur between leaf layer and soil, layered canopies, and rough soil [4].

### 3. ALGORITHM TAXONOMIES: LINEAR UNMIXING

We return to the linear mixing model in (1). In its most complete form, linear unmixing is a sequence of three consecutive steps that estimates the spectra of the constituent substances, or endmembers, and the associated fractional abundances for each pixel in the scene:

**Dimension reduction** Reduce the dimension of the data in the scene. This step is optional and is only invoked by some algorithms in order to reduce the computational load of subsequent steps.

**Endmember determination** Estimate the distinct spectra that constitute the mixed pixels in the scene.

**Inversion** Estimate the fractional abundances of each mixed pixel from its spectrum and the endmember spectra.

Most of the algorithms discussed in this document address only one of these steps. In some cases, however, an algorithm may estimate endmembers and abundances concurrently, effectively performing two steps simultaneously.

#### 3.1 TAXONOMY STRUCTURE

In creating the taxonomies in Figures 3, 4, and 5, we have attempted to both group together as well as differentiate algorithms by their characteristics. From extensive compilations of their properties, algorithms were then hierarchically categorized by three features, depicted in Figure 2, that highlight important philosophical interpretations of the unmixing problem. Figures 3, 4, and 5 all share the top-down hierarchical structure, whose features and categories are defined as follows:

**Interpretation of data** Indicates how an algorithm interprets mixed pixel spectra.

- **Non-statistical vs. Statistical:** If an algorithm employs statistical measures (e.g., means, covariances) to quantify the aggregate behavior of hyperspectral data, it is statistical. Implicitly, statistical algorithms process mixed pixels by comparing their features to those belonging to a known class. This distinction becomes important in applications like detection where aggregate measures can obscure the features of rare targets.

**Description of randomness** Indicates how an algorithm incorporates the randomness of the data.

- **Non-parametric vs. Parametric:** If an algorithm presumes the random behavior of hyperspectral data is described by a specific, parameterized probability distribution function, it is parametric. An algorithm that is statistical is not always parametric, although the converse must be true. For example, algorithms that empirically estimate means and covariances without assuming a specific model for the underlying density are statistical, but non-parametric.

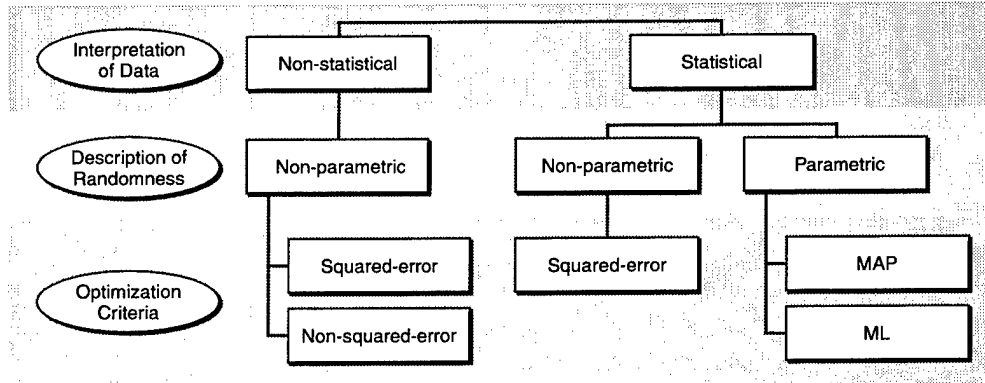


Figure 2. Conceptual structure for taxonomies.

**Optimization criteria** Indicates what objective function is being optimized.

- **Non-squared-error and squared-error:** If an algorithm employs squared-error as a distance measure, it is classified as using squared-error.
- **Maximum Likelihood (ML):** Parametric algorithms that estimate deterministic quantities by maximizing a likelihood function derived from a forward density are considered ML.
- **Maximum a Posteriori (MAP):** Parametric algorithms that estimate random quantities by maximizing a likelihood function derived from a posterior density are considered MAP.

Further discrimination in the taxonomies beneath the structure in Figure 2 is accomplished by examining features pertaining to 1) the primary output of the algorithm, 2) the required inputs, and 3) and the characterization of noise. More detailed information for each taxonomy appears in its respective section.

Finally, the taxonomies enumerate the remaining features of unmixing algorithms that are not incorporated into the hierarchical, taxonomy structure. They have been coarsely grouped into categories related to 1) a priori information required as inputs for an algorithm, 2) output features, 3) additional algorithm outputs besides the intended result, and 4) processing and architectural requirements.

### 3.2 DIMENSION REDUCTION TAXONOMY

In this section, we consider techniques to reduce the dimension of hyperspectral data. The taxonomy for this family of algorithms appears in Figure 3. In addition to the three levels of structure imposed by the features in Figure 2, three more features provide further discrimination for dimension reduction algorithms: 1) the output axes for compression, 2) the required input data, and 3) the kind of noise, if any, in the signal model.

347586-2

Interpretation of data		Non-statistical	Statistical
Description of Randomness		Non-parametric	
Specifics	Optimization Criterion	Squared-error	Non-squared-error
	Outputs	Orthogonal axes	SNR-based
	Inputs	Utilizes reduced data set	Non-orthogonal axes
	Noise	No noise model	Additive obs. noise
Algorithm		ORASIS	Principal Components Analysis
Required A Priori Info.	Prescreen Threshold	Yes	No
	Noise Covariance	No	Yes
Processing	Iterative	No	
	Block-pixel Processing	No	Yes
	Real-time	Yes	No
	Requires Inverse	No	Yes
Requires PCA/SVD		No	Yes
		Maximum Noise Fraction/ Noise Adjusted Principal Components	

Figure 3. Taxonomy of dimension reduction algorithms.

The simplest form of dimension reduction discards unwanted bands altogether. This approach is useful when the utility of specific bands is well known, or, conversely, when irrelevant or unreliable bands are identifiable. Such decisions are usually justified by applications demanding strong physical knowledge. In lieu of such knowledge, the dimension reduction algorithms discussed here attempt to compress hyperspectral data based on alternative criteria.

### 3.2.1 Non-statistical and Non-parametric Dimension Reduction

Non-statistical techniques strive to compress hyperspectral data by identifying the useful components of pixel spectra, but they avoid using aggregate measurements of the data (e.g., means, covariances) as the basis for calculations. Instead, they rely on “growing” and “pruning” the useful subspace from a stream of consecutive pixels. While optimality may be lost in one context, these methods are often better suited to real-time implementation because they avoid the latency intrinsic to statistical methods.

**3.2.1.1 Squared-error** In this section we consider one algorithm that employs a measure of squared-error to determine the axes for compression of pixel spectra.

**ORASIS** A non-statistical technique for reducing the dimension of hyperspectral data appears in the Naval Research Laboratory’s ORASIS (Optical Real-time Adaptive Spectral Identification System), which is a series of hyperspectral processing modules [5, 6]. In the Exemplar Selector Module (ESM), when a new pixel is collected from the scene by the sensor, its spectra is compared to the existing set of exemplars (the first pixel in a scene automatically becomes the first exemplar). If the new pixel is sufficiently different – the usual measure of distance is spectral angle – from each of the existing exemplars based on a prescreen threshold, it is added to the collection. If it is not sufficiently different, the exemplar set remains unchanged. At periodic intervals, the Salient Selection Module (SSM) then orthogonalizes the current set of exemplars using a modified Gram-Schmidt process where a permitted representation threshold, determines what subset of the orthogonal vectors are retained to reduce the dimension of the current set of exemplars. In this reduced dimension, the exemplars are submitted to another module to geometrically determine endmembers.

The performance of the algorithm is linked to the value of the prescreen angle threshold which is used to admit exemplars and to the representation threshold that defines the dimension of the projected space. Smaller values of the prescreen threshold admit more exemplars, and generally, add more dimensions to the exemplar subspace. This parameter may be tuned to admit different levels of variability in the scene into the exemplar set. The representation threshold, however, defines the required dimension of the projected space to preserve a minimum fidelity in the exemplars and is important to the subsequent module for determining endmembers

### 3.2.2 Statistical and Non-parametric Dimension Reduction

In this section we present statistical techniques for reducing the dimension of hyperspectral data that possess no parametric models. The axes for compression in all of these algorithms are



revealed by performing an eigendecomposition on covariance matrices. Features of targets that occur with a relatively low probability may be lost using statistical compression techniques, and the requisite processing may be unsuitable for real-time implementations. However, the optimal properties associated with statistical methods make them desirable.

**3.2.2.1 Squared-error** We consider a technique for reducing the dimension of pixels in an image that directly exploits the squared-error approximation properties of the eigendecomposition [31].

**Principal Component Analysis** Principal component analysis (PCA) performs an eigendecomposition on the covariance of the data and determines how the power in a signal is allocated. Letting  $\mathbf{x}(n)$ ,  $n = 1, \dots, N$ , denote the spectra of  $N$  mixed pixels, the covariance,  $\mathbf{\Gamma}_x$ , is calculated as

$$\mathbf{\Gamma}_x = E[(\mathbf{x} - \mu_x)(\mathbf{x} - \mu_x)^T] = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}(n) - \mu_x)(\mathbf{x}(n) - \mu_x)^T \quad (3)$$

where  $\mu_x$  is the mean vector of the pixel set. The resulting eigendecomposition can be expressed as  $\mathbf{\Gamma}_x = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$ , where  $\mathbf{U}$  is a unitary matrix of eigenvectors, and  $\mathbf{\Sigma}$  is a diagonal matrix of eigenvalues. The magnitude of the eigenvalues indicates the power residing in the data along the component of the data parallel to the associated eigenvector. The larger eigenvalues identify basis components whose absence in  $\mathbf{x}$  is directly proportional to a larger average squared-error. Hence, the effective dimensionality of the data can be estimated by counting the number of non-zero eigenvalues. If the data is transformed by pre-multiplication with  $\mathbf{U}^T$ , then the resulting linear transformation moves  $\mathbf{a}$  to a new system of decorrelated variables oriented along the eigenvectors in  $\mathbf{U}$ . The new vector of variables can be truncated to retain only those having significant eigenvalues. The result is a lower-dimensional multivariate random vector that conveys most of the power in the original, higher-dimensional system.

**3.2.2.2 Signal-to-Noise Ratio** We examine an alternative technique that maximizes signal-to-noise ratio rather than minimizing squared-error, but still employs the eigendecomposition as its principle computational tool.

**Maximum Noise Fraction/Noise Adjusted Principal Components** One technique for compression that is also based on the eigendecomposition produces an ordered decomposition of a signal into components based on the ratio of noise power to signal power [12]. Provided the second-order statistics of the noise in (1) are known, let the noise and signal and noise be uncorrelated and known by  $\mathbf{S}\mathbf{\Gamma}_a\mathbf{S}^T$  and  $\mathbf{\Gamma}_w$ , respectively, such that  $\mathbf{x}$  has a covariance given by  $\mathbf{\Gamma}_x = \mathbf{S}\mathbf{\Gamma}_a\mathbf{S}^T + \mathbf{\Gamma}_w$ . The component of the received signal,  $\mathbf{v}$ , possessing maximum noise fraction maximizes

$$\frac{\mathbf{v}^T \mathbf{\Gamma}_w \mathbf{v}}{\mathbf{v}^T \mathbf{\Gamma}_x \mathbf{v}}. \quad (4)$$

The maximum noise fraction (MNF) transform is a linear transformation that identifies and orders signal components by their noise fraction. The eigenvalues of  $\Gamma_w \Gamma_x^{-1}$  are the noise fractions associated with each left-hand eigenvector. Consequently the left-hand eigenvectors of  $\Gamma_w \Gamma_x^{-1}$  comprise a linear, orthogonal transform that orders the components of a signal by their noise fractions. This is in direct contrast to PCA where no noise model is used, and the components are ordered only by their signal power. In MNF, the largest eigenvalues identify components of the signal exhibiting the highest noise fractions, while smaller eigenvalues identify the opposite. It is important to note that the noise fraction says nothing about that the absolute signal or noise levels. This information is obtained from the PCA of the individual covariance matrices.

The Noise Adjusted Principal Components (NAPC) transform [20] orders components in a signal by decreasing signal-to-noise ratio (SNR). Aside from the ordering of the components, the transform is mathematically equivalent to the MNF transform and achieves the exact same result. As in PCA, the ordering of components can be used to estimate one type of effective signal dimensionality, and the set of random variables obtained after the MNF transform can be truncated to retain only those that have a useful SNR.

### 3.3 ENDMEMBER DETERMINATION TAXONOMY

In this section, we examine different endmember determination algorithms. The taxonomy in Figure 4 organizes the algorithms along the lines of the features in Figure 2. To add further structure, four additional features have been included: 1) the type of endmembers being derived, 2) the number of endmembers,  $P$ , estimated by an algorithm, with respect to the number of spectral bands,  $M$ , and the number of pixels being processed,  $N$ , 3) the required input data, and 4) the kind of noise, if any, in the signal model.

#### 3.3.1 Non-statistical and Non-parametric Endmember Determination

In this section we consider two ways to geometrically interpret hyperspectral data. Both methods derive endmembers from a scene without statistical measures or parametric models.

**3.3.1.1 Geometric Approaches** A purely geometric approach toward endmember determination has been explored to exploit the strong parallelism between the linear mixing model and the theory of convex sets [3, 9]. Hyperspectral data from a scene reside within a volume in a multi-dimension hyperspace; endmembers are estimated from the most extreme-valued pixels in the scene, so that any point within the volume is a non-negative sum of those endmembers.

The determination of endmembers consists of two steps. The first requires finding a minimum subset of pixels in the scene that describe the range of variation in the scene. The convex hull of the pixels serves this purpose, and is always unique for every scene, although various methods exist for its calculation [26]. The second step involves fitting a simplex around the convex hull that satisfies a criteria for ‘best fit,’ and unlike the convex hull, the simplex is not unique. The simplex is a geometric surface in an  $n$ -dimensional hyperspace defined by exactly  $n + 1$  vertices. A simplex enclosing the convex hull of pixels in a scene can be arbitrarily large, but if the enclosed volume is

Interpretation of data		Non-statistical			Statistical														
Description of Randomness		Non-parametric										Parametric							
Optimization Criterion	Non-squared-error			Squared-error							ML								
	Geometric																		
Specifics	Outputs	P = M			P = M+1			P = M			P = M			Stochastic (Gaussian) Endmembers			Non-Gaussian Deterministic Endmembers		
		Convex hull of pixel spectra			All pixel spectra and abundances			All pixel spectra and abundances			All pixel spectra			All pixel spectra					
	Inputs	No noise model			Additive meas. noise			Additive obs. noise			Gaussian			No noise model					
Algorithm	Minimum Volume Transform	Penalty-based Shrinkwrapping		Block Least Squares		Total Block Least Squares		Fuzzy K-Means		Non-linear Least Squares		Gaussian Class Estimation		Independent Component Analysis					
		DPFT	FPFT																
Dark Point		Yes		No															
Required A Priori Info.	Noise Statistics			No				Yes		No									
	Reference Endmember Spectra			No						Yes									
Additional Outputs	Priors for Endmembers			No				Yes		No									
	Abundances			No				Yes		No									
Processing	Iterative	Yes		No				Yes											
	Block-pixel Processing	No		Yes				No		Yes		No							
	Requires Inverse	No		Yes				No		Yes		No							
	Requires SVD/PCA	No		Yes				No											
Uses Neural Networks				No								Yes							

Figure 4. Taxonomy of endmember determination algorithms.

minimized by adjusting its faces, its vertices can be interpreted as the locations of pure spectra in the scene and may be used as endmembers in inversion.

Two different algorithms for estimating endmembers geometrically are derived from an iterative technique for encapsulating the data cloud with a simplex having minimum volume [9]. They are distinguished by a priori knowledge of a dark point that is defined as the response of the sensor to a target of zero reflectance.

**Minimum Volume Transform: Dark-Point-Fixed Transform** The dark-point-fixed transform (DPFT) [9] is presented for the scenario where the dark point is known and the co-ordinate axes can be adjusted to that point; put differently, the origin is permitted to be one vertex of the simplex. First, the hyperplane,  $H = \{\mathbf{x}_H \in \mathbb{R}^M : \mathbf{u}^T \mathbf{x}_H = 1\}$ , is introduced, where  $\mathbf{u}$  is an  $M \times 1$  vector of all ones whose length equals the number of bands, and  $\mathbf{x}_H$  is the normalized projection of a pixel,  $\mathbf{x}$ , lying in  $H$ . The projection of a set of pixels on  $H$  identifies a region to be enclosed by a new set of co-ordinate vectors, also residing on  $H$ , that encapsulate the entire set of points, but with a minimum volume. If  $\mathbf{S}$  is an  $M \times M$  matrix whose columns,  $\mathbf{s}_k$ , also reside on  $H$ , and  $\mathbf{Q} = \mathbf{S}^{-1}$ , then  $\mathbf{u}^T \mathbf{S} = \mathbf{u}^T$  and  $\mathbf{u}^T \mathbf{Q} = \mathbf{u}^T$ . The volume of the simplex,  $S$ , formed by the columns of  $\mathbf{S}$  and capped by  $H$  is

$$\text{volume}(S) = \frac{\text{abs}(|\mathbf{S}|)}{M!}. \quad (5)$$

If  $\mathbf{x}_H$  represents the value of pixels in the standard, orthonormal axes, then, in the co-ordinate system of the non-orthogonal columns of  $\mathbf{S}$ , the equivalent co-ordinate values are the mixing coefficients,  $\mathbf{a}_H = \mathbf{S}^{-1} \mathbf{x}_H = \mathbf{Q} \mathbf{x}_H$ .

The goal is to minimize the volume of  $S$ , but by (5), we can alternately maximize  $|\mathbf{Q}|$  by iteratively adjusting each of its rows. This is equivalent to adjusting one facet of the simplex by manipulating its surface normal while keeping the others fixed. This technique for maximizing  $|\mathbf{Q}|$  is shown to be equivalent to maximizing the product of a constrained set of variables, and numerical approaches from linear programming are used [9]. However, in the more general situation, the dark point is unknown, and the minimum volume simplex, defined by  $\mathbf{S}$ , can no longer assume the origin is an endmember, requiring  $\mathbf{S}$  to explicitly contain an additional column. Since the number of vertices in a simplex exceeds the dimension of the data,  $M$ , by one, the new matrix of endmembers,  $\mathbf{S}$ , is now  $M \times (M + 1)$ .

**Minimum Volume Transform: Fixed-Point-Free Transform** By contrast, in the fixed-point-free transform (FPFT) [9], the additional endmember is accommodated in  $\mathbf{S}$  by adding a row of ones to  $\mathbf{S}$ , making it an  $(M + 1) \times (M + 1)$  matrix,  $\mathbf{S}_{aug}$ . The implication of  $\mathbf{S}_{aug}$  is that the simplex,  $S$ , residing in  $\mathbb{R}^M$  is embedded within a simplex,  $T$ , in  $\mathbb{R}^{M+1}$  having one vertex at the origin and the others at co-ordinates represented by the columns of  $\mathbf{S}_{aug}$ . The volume associated with  $T$  is given by

$$\text{volume}(T) = \frac{1}{M + 1} \text{volume}(S) = \frac{1}{M(M + 1)} \frac{\text{abs}(|\mathbf{S}_{aug}|)}{M!} \quad (6)$$

where (5) has been invoked. So, minimizing  $|S_{aug}|$  equivalently minimizes the volume of  $S$  using the theory of determinants. The same method of individually modifying each of the surface normals is employed for the FPFT where now  $\mathbf{Q}_{aug} = \mathbf{S}_{aug}^{-1}$ . The  $M + 1$  endmembers for the simplex of interest in  $\mathbb{R}^M$  can be obtained from the columns of the final version of  $\mathbf{S}_{aug}$  by retaining only the first  $M$  rows.

**Penalty-Based Shrinkwrapping** The DPFT and FPFT build simplices having minimum volume. Some pixels in the scene, however, may reside very close to the facets of the simplex, causing potential instability during inversion. One method for avoiding this scenario strives to build a simplex having minimum volume, but is constrained by a penalty function that discourages facets from residing too close to data points [11].

Adopting block notation for  $N$  pixels, we let  $\mathbf{S}$  be the  $M \times (M + 1)$  matrix whose columns are endmembers.  $\mathbf{S}$ , however, is not square, and, hence, is not invertible. The fractional abundances of each pixel are constrained to be positive and must sum to 1. This latter constraint can be appended to the block formulation of (1), yielding

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{1}^T \end{bmatrix} \quad \tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{S} \\ \mathbf{1}^T \end{bmatrix} \quad \tilde{\mathbf{E}} = \begin{bmatrix} \mathbf{W} \\ \mathbf{0}^T \end{bmatrix} \quad (7)$$

where  $\mathbf{1}$  is a column vector of ones, and  $\mathbf{0}$  is a column vector of zeroes. The synthesis equation can be expressed as  $\tilde{\mathbf{X}} = \tilde{\mathbf{S}}\mathbf{A} + \tilde{\mathbf{W}}$ . An estimate of  $\mathbf{A}$  can be recovered by a simple inversion,  $\hat{\mathbf{A}} = \tilde{\mathbf{S}}^{-1}\tilde{\mathbf{X}}$ .

The objective function,  $H$ , involves two terms, a volume term and a penalty term, and is defined as

$$H(\mathbf{S}, \mathbf{X}) = V(\mathbf{S}) + \alpha F(\mathbf{S}, \mathbf{X}). \quad (8)$$

$V(\mathbf{S})$  is the volume of the simplex, calculated using (5) and (6).  $F(\mathbf{S}, \mathbf{X})$  is a penalty function enforcing the constraint that all data points,  $\mathbf{X}$ , reside inside the simplex. The parameter,  $\alpha$ , weighs the impact of the penalty on  $H(\mathbf{S}, \mathbf{X})$ . Intuitively,  $F(\mathbf{S}, \mathbf{X})$  should be small when the facets of the simplex are far from the points in  $\mathbf{X}$ , and large when the opposite is true.

The penalty function is chosen as

$$F(\mathbf{S}, \mathbf{X}) = \sum_{i=1}^P \sum_{j=1}^N \frac{1}{\hat{A}_{ij}}. \quad (9)$$

In order to determine a value of  $\mathbf{S}$  that minimizes  $H(\mathbf{S}, \mathbf{X})$ , the gradient of (8) with respect to the columns of  $\mathbf{S}$  is set to zero and solved for a recursion that leads to a final estimate for  $\mathbf{S}$  that minimizes  $H(\mathbf{S}, \mathbf{X})$ .

### 3.3.2 Statistical and Non-parametric Endmember Determination

In this section, we consider algorithms that do not assume a parametric model for randomness, but still use statistical information to estimate endmembers.

**3.3.2.1 Squared-error** In this section, we consider non-parametric algorithms that estimate endmembers in a scene grouped loosely around the concept of minimizing a squared-error criterion. The discussion of the least squares problem is well-documented [19]. Although other techniques, especially those involving Gaussian distributions, ultimately minimize a quadratic metric, we defer the discussion of parametric models to later sections.

**Block Least Squares** We consider the problem of estimating the endmembers in a scene when both spectral observations and ground class abundance measurements are collected. In essence, this is the case of having  $\mathbf{x}$  (measurements) and  $\mathbf{a}$  (ground truth) for a set of pixels, and estimating  $\mathbf{S}$  in (1). Depending on how the sources of error are modeled, two different least squares estimators can be derived [15]. We use (1) as a starting point and adopt block notation for  $N$  pixels, with the goal of developing an estimate for  $\mathbf{S}$  that minimizes  $\|\mathbf{X} - \hat{\mathbf{S}}\mathbf{A}\|_F^2$ . Here,  $\|\cdot\|_F$  is the Froebenius norm [31].

We develop a least squares error estimate for  $\mathbf{S}$  in light of the fact that  $\mathbf{X} = \mathbf{X}_{exact} + \mathbf{X}_{error}$  and  $\mathbf{A} = \mathbf{A}_{exact} + \mathbf{A}_{error}$ . Errors in both values may occur for a number of reasons.  $\mathbf{X}_{error}$  maybe an instrumentation error corrupting the observation in  $\mathbf{X}$  and  $\mathbf{A}_{error}$  may be the human error in assessing ground class proportions distorting the measurement in  $\mathbf{A}$ . We now explore different estimates for  $\mathbf{S}$  based on the assumption that one or both errors are present and assuming that  $\mathbf{X}_{exact} = \mathbf{S}\mathbf{A}_{exact}$ .

First, we generate a least squares error estimate of  $\mathbf{S}$  based on the assumption that  $\mathbf{X}$  possesses only an observation error, in which case  $\mathbf{X} = \mathbf{X}_{exact} + \mathbf{X}_{error} = \mathbf{S}\mathbf{A}_{exact} + \mathbf{W}$ . We can recover an estimate for  $\mathbf{S}$  by multiplying both sides by the pseudo-inverse [22] of  $\mathbf{A}$ , defined as  $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ . This leads to the block least squares estimate for the endmembers,  $\hat{\mathbf{S}}_{BLS} = \mathbf{X}\mathbf{A}^+$ .

**Total Block Least Squares** If both  $\mathbf{A}$  and  $\mathbf{X}$  are assumed to be noisy, then  $\mathbf{X} = \mathbf{X}_{exact} + \mathbf{X}_{error} = \mathbf{S}\mathbf{A}_{exact} + \mathbf{S}\mathbf{A}_{error} + \mathbf{W}$ . We can extend the analysis for  $\mathbf{S}_{BLS}$  to obtain a block total least squares estimate,  $\mathbf{S}_{BTLS}$ , which accounts for the measurement error in  $\mathbf{A}$ . We can rewrite (1) as

$$[\mathbf{A}^T \mathbf{X}^T][\mathbf{S} \quad -\mathbf{I}]^T = \mathbf{C}[\mathbf{S} \quad -\mathbf{I}]^T = \mathbf{0} \quad (10)$$

where  $\mathbf{C} = [\mathbf{A}^T \mathbf{X}^T]$  is an  $N \times (M + P)$  matrix whose rows are the concatenation of  $N$  abundance measurements with their associated pixel spectrum observations.

Letting the singular-value decomposition (SVD) [31] of  $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ,  $\mathbf{V}$  may be partitioned as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \quad (11)$$

where  $\mathbf{V}_{11}$  is  $P \times P$  and  $\mathbf{V}_{22}$  is  $M \times M$ . It can be shown [15] that the total block least squares solution for  $\hat{\mathbf{S}}$  is

$$\hat{\mathbf{S}}_{BTLS} = -(\mathbf{V}_{12}\mathbf{V}_{22}^{-1})^T. \quad (12)$$

**Fuzzy K-Means Partitions** Clustering algorithms attempt to identify natural partitions in data that exemplify distinct statistical behavior. This method has been adapted to estimate the endmembers and abundances from data in a scene so that a global metric based on squared-error is minimized [10]. For a collection of  $N$  pixels, the objective function may be expressed as

$$J_m(\mathbf{A}, \mathbf{S}) = \sum_{i=1}^P \sum_{j=1}^N (A_{ij})^m (d_{ij})^2 \quad (13)$$

where  $A_{ij}$  is an estimate of the  $i$ -th abundance for the  $j$ -th pixel and  $d_{ij}$ , the weighted squared-error between the  $j$ -th pixel and the  $i$ -th centroid, or endmember,  $\mathbf{s}_i$ , is

$$(d_{ij})^2 = (\mathbf{x}_j - \mathbf{s}_i)^T \mathbf{W} (\mathbf{x}_j - \mathbf{s}_i). \quad (14)$$

The weighting matrix,  $\mathbf{W}$ , permits certain abundances to be weighted more than others. The minimization of  $J_m(\mathbf{A}, \mathbf{S})$  is constrained by physical restrictions on the values of abundances (see Section 3.4). These requirements are inserted into the estimator for  $\mathbf{A}$ . As noted,  $m$  indicates the degree of fuzziness, and the shapes of the regions associated with each class centroid vary as  $m$  is changed. Suggested values for  $m$  fall in the range,  $1.5 < m < 3.0$ . The optimization of  $J_m(\mathbf{A}, \mathbf{S})$  is accomplished iteratively using a variant of the  $K$ -means clustering algorithm [2], until  $J_m(\mathbf{A}, \mathbf{S})$  achieves a minimum using a final estimate of the endmembers,  $\hat{\mathbf{S}}$ , and abundances,  $\hat{\mathbf{A}}$ .

### 3.3.3 Statistical and Parametric Endmember Determination

In this section, we consider approaches to determining endmembers that utilize statistics and assume an underlying, parametric model for the received pixel spectra.

**3.3.3.1 Maximum Likelihood Estimation** To derive a maximum likelihood (ML) estimate of a deterministic variable, a forward density for the received data is maximized. In this section we explore ML algorithms for estimating endmembers. However, while these algorithms estimate non-random parameters, this does not necessarily mean the endmembers themselves are deterministic. Stochastic endmembers generalize the concept of deterministic spectra to random spectra having parametric, multivariate representations. So even if the endmembers are stochastic, an ML algorithm can still estimate the deterministic parameters of the endmember densities.

**Non-linear Least Squares** The unmixing problem can be formulated as an inverse problem [32] whose objective is to estimate system parameters from observations. By virtue of the assumptions of linear mixing and Gaussianity, the problem can be reduced to that of ML estimation using a quadratic objective function containing the unknown, deterministic, system parameters. However, when both endmembers and abundances are being estimated simultaneously in the linear mixing model, the minimization of quadratic error results in a non-linear least squares formulation. Starting with the general formulation [32], this concept has been adapted to an iterative algorithm that is initialized by an a priori estimate for endmembers and abundances and simultaneously derives an ML estimate for endmembers and fractional abundances [33].

The elements of  $\mathbf{S}$  and  $\mathbf{a}$  are unknown, deterministic parameters that we combine in a single  $(MP + P) \times 1$  Gaussian random vector,  $\mathbf{g} = [S_{11}, \dots, S_{MP}, \mathbf{a}^T]^T$ . Then, the following recursion for  $\hat{\mathbf{g}}$  is given by

$$\hat{\mathbf{g}}^{(k+1)} = \hat{\mathbf{g}}^{(k)} + [(\mathbf{G}^{(k)})^T \mathbf{\Gamma}_w^{-1} \mathbf{G}^{(k)} + \mathbf{\Gamma}_{g^{(0)}g^{(0)}}^{-1}]^{-1} [(\mathbf{G}^{(k)})^T \mathbf{\Gamma}_w^{-1} (\mathbf{x} - \hat{\mathbf{S}}^{(k)} \hat{\mathbf{a}}^{(k)}) - \mathbf{\Gamma}_{g^{(0)}g^{(0)}}^{-1} (\hat{\mathbf{g}}^{(k)} - \mathbf{g}^{(0)})]. \quad (15)$$

$\hat{\mathbf{g}}^{(k)}$  is the  $k$ -th estimate of  $\mathbf{g}$  which contains  $\hat{\mathbf{S}}^{(k)}$  and  $\hat{\mathbf{a}}^{(k)}$ .  $\mathbf{g}^{(0)}$  is the initial guess for  $\mathbf{g}$ , and  $\mathbf{\Gamma}_{g^{(0)}g^{(0)}}$  is its covariance matrix under the assumption that  $\mathbf{g}^{(0)}$  is Gaussian.  $\mathbf{x}$  is the received pixel spectra having  $\mathbf{\Gamma}_w$  as its covariance which is equal to the additive noise covariance.  $\mathbf{G}^{(k)}$  is the constraint matrix at the  $k$ -th iteration.

A priori estimates of mixing coefficients and endmembers can originate from previous experiments or from laboratory measurement, and the degree of confidence in these initial estimates is specified in their respective covariance estimates which comprise  $\mathbf{\Gamma}_{g^{(0)}g^{(0)}}$ . All pixels in an image are modeled simultaneously, and consequently, the number of mixing coefficients being solved for is the number of total pixels multiplied by  $P$ . The iterative algorithm in (15) is repeated until the difference between successive estimates falls below a prescribed threshold. Clearly,  $\hat{\mathbf{g}}^{(k+1)}$  consists of an adjustment to  $\hat{\mathbf{g}}^{(k)}$  based on the weighted sum of the parameter vector error,  $(\hat{\mathbf{g}}^{(k)} - \mathbf{g}^{(0)})$ , and the spectrum error,  $(\mathbf{x} - \hat{\mathbf{S}}^{(k)} \hat{\mathbf{x}}^{(k)})$ . In the case of both errors, the difference is measured with reference to the starting values in  $\mathbf{g}^{(0)}$ , not the previous value of  $\hat{\mathbf{g}}$ . The implication here is that dramatic changes in estimates, when gauged against the values in  $\hat{\mathbf{g}}$ , are inversely weighted by the associated confidence in the original estimates appearing in  $\mathbf{\Gamma}_{g^{(0)}g^{(0)}}$ .

**Gaussian Class Estimation** An adaptation of the linear mixing model combines the geometric approach in Section 3.3.1.1 with stochastic data clustering approaches [30]. The stochastic mixing model (SMM) introduces the concept of hard endmember classes as the stochastic extension of deterministic endmembers and assumes that all data in a scene is Gaussian and arises from a linear combination of at least one hard class. Geometrically, each stochastic endmember describes a cluster of pixels on the perimeter of the data cloud, and each pixel should belong to a particular combination of hard classes. If each fractional combination of hard classes is a mixed class, with its own Gaussian statistics – mean, covariance, and prior probability – then a pixel may be unmixed by determining which combination of classes has the greatest likelihood of realizing that pixel. After an initial partition of the data, the newly re-classified pixels are used to generate new class statistics until the associated SMM class parameters no longer change. This procedure is the stochastic equivalent of the expectation maximization (EM) algorithm for ML estimation [23].

We can demonstrate this concept for two hard classes. Let a scene be comprised of two hard classes,  $\mathbf{x}_1 \sim N(\mu_1, \mathbf{\Gamma}_1)$  and  $\mathbf{x}_2 \sim N(\mu_2, \mathbf{\Gamma}_2)$ , each having a mean and covariance that describes the  $M$ -dimensional Gaussian density function for pixels belonging to that class. Then, a mixed pixel is a linear combination of observations arising from the two hard classes so that

$$\mathbf{x} = f\mathbf{x}_1 + (1 - f)\mathbf{x}_2 \quad (16)$$

and, consequently, for a fixed value of  $f$ , the corresponding statistics are  $\mu_f = f\mu_1 + (1 - f)\mu_2$  and  $\mathbf{\Gamma}_f = f^2\mathbf{\Gamma}_1 + (1 - f)^2\mathbf{\Gamma}_2$  when  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are uncorrelated. Since  $f$  lies in the interval,  $[0, 1]$ , an



infinite number of soft classes exists, but if  $f$  is discretized on  $[0, 1]$ , a finite number of soft classes is attained, each described by the resulting Gaussian statistics. If the number of hard classes,  $C_h$ , is greater than 2,  $f$  must be a vector of length  $C_h - 1$ , whose elements are discretized, describing the contribution from the  $C_h$  hard classes. If  $C$  is the total number of hard and mixed classes, then each class,  $c = 1, \dots, C$ , has a mean, covariance, and prior probability,  $(\mu_c, \Gamma_c, \Pi_c)$ .

The number of hard classes,  $C_h$ , is manually estimated before iterating by matching pixel regions in the scene with reference spectra. Given a discretization factor for the fractional abundance of each hard class, a lattice of acceptable soft classes is created to represent the space of all possible mixing coefficients. The iterative estimation of class parameters  $(\mu_c, \Gamma_c, \Pi_c)$  is a maximum likelihood (ML) search of class statistics that is performed by employing a stochastic variation on the traditional Expectation-Maximization (EM) algorithm [23]. The class parameters are refined after each round of reclassification until successive iterations yield negligible changes, and the abundances for each pixel are ascertained from the mixture describing the nearest class.

**Independent Component Analysis** The linear mixing model predicts that the signal received by a sensor from a scene is the sum of spectra from distinct substances. The substances may be man-made or natural, and without prior knowledge, the challenge of unmixing is to identify the spectra of the components and the associated abundances. A key assumption is that the spectra are statistically independent of each other with the physical significance being that the presence of one substance in a mixed pixel should not interfere with the scattering of another, regardless of the quantities in which they appear. This is a significantly stronger condition than the more traditional objective of decorrelation present in Gaussian analysis.

Beyond decorrelation, independence means that the parametric densities for the spectra of substances are separable. If  $p_{\mathbf{x}_1}(\mathbf{x}_1)$  and  $p_{\mathbf{x}_2}(\mathbf{x}_2)$  are the densities from two distinct substances, then independence implies

$$p_{\mathbf{x}_1, \mathbf{x}_2}(\mathbf{x}_1, \mathbf{x}_2) = p_{\mathbf{x}_1}(\mathbf{x}_1)p_{\mathbf{x}_2}(\mathbf{x}_2). \quad (17)$$

Independent component analysis (ICA) is recognized as a special case of the blind source separation problem where the objective is to estimate the parameters of statistically independent sources,  $\mathbf{s}$ , that combine through a mixing matrix,  $\mathbf{W}$ , to form an array of observations,  $\mathbf{x}$ , collected at the sensor. Considered as a time series, each element of the array possesses its own statistics, and under the assumption of linear mixing, they can be used to estimate the mixing coefficients and the parameters of the source distributions so that the statistical independence of the sources is maximized.

In the adaptation of ICA to spectral unmixing, however, we assume endmembers are deterministic; but, we can consider a collection of similar mixed pixels to be multiple occurrences of mixing by the same independent sources, or endmembers, that obey  $\mathbf{x} = \mathbf{W}\mathbf{s}$ . In lieu of a time series to derive statistics, the spectral values for each mixed pixel are converted to a histogram. Likewise, a similarly synthetic histogram for each of the sources is posited from the unknown end-member spectra. Based on an assumed parametric form for the cumulative distribution functions for these synthetic densities, ICA estimates the shape of the source distribution functions that

maximize the independence of the endmember histograms, and still combine linearly to yield the received histograms. In order to explicitly identify the endmembers in the scene, these histograms are then correlated with distribution functions derived from histograms of reference spectra [1].

To maximize the independence of the sources, the Kullback-Leibler distance,  $G$ , compares the synthetic histogram observed at the sensor with the one derived from the estimated linear mixture of independent sources.

$$G(p_{\mathbf{x}}(\mathbf{x}), \hat{p}_{\mathbf{x}}(\mathbf{x}; \mathbf{w}, \mathbf{W})) = \int p_{\mathbf{x}}(\mathbf{x}) \log \frac{p_{\mathbf{x}}(\mathbf{x})}{\hat{p}_{\mathbf{x}}(\mathbf{x}; \mathbf{w}, \mathbf{W})} d\mathbf{x} = H(\mathbf{x}) - \int p_{\mathbf{x}}(\mathbf{x}) \log \hat{p}_{\mathbf{x}}(\mathbf{x}; \mathbf{w}, \mathbf{W}) d\mathbf{x}. \quad (18)$$

No additive noise is considered in this model. Here,  $p_{\mathbf{x}}(\mathbf{x})$  is the synthetic histogram for  $\mathbf{x}$  that is manufactured empirically from the received spectra. The histogram for the linear model estimate of  $\mathbf{x}$  is  $\hat{p}_{\mathbf{x}}(\mathbf{x}; \mathbf{w})$ , where  $\mathbf{w}$  is a vector containing the parameters defining the densities of the sources as well as the mixing coefficients,  $\mathbf{W}$ . To minimize (18), partial derivatives with respect to the elements of  $\mathbf{w}$  yield a learning rule for a neural network to arrive at an ML estimate for  $\mathbf{S}$ .

### 3.4 INVERSION TAXONOMY

In this section we examine numerous algorithms that yield fractional abundances,  $\hat{\mathbf{a}}$ , from a mixed pixel spectrum,  $\mathbf{x}$ . Figure 5 illustrates the taxonomy of these inversion algorithms. Inversion is essentially one form of the classical parameter estimation problem [35]. Given a vector of observations,  $\mathbf{x}$ , and a signal model, we wish to estimate a vector of parameters,  $\mathbf{a}$ . As in Sections 3.2 and 3.3, we organize the techniques in this section according to the features in Figure 2. To add further structure, two specific features have been included: 1) the type of values estimated abundances may have, 2) the type of endmembers required, and 3) the kind of noise, if any, in the signal model.

Any meaningful estimate of  $\mathbf{a}$ , however, must comply with constraints that make it physically realizable. In fact, the single most challenging aspect of unmixing is determining how to reconcile mathematical and statistical techniques with the underlying physical restrictions. With this in mind, any estimate of  $\mathbf{a}$  must obey the following constraints:

**Non-negativity** The abundances should be non-negative to be meaningful in a physical sense:

$$a_i \geq 0, i = 1, \dots, P.$$

**Purity** A fractional abundance coefficient should not exceed 1:  $a_i \leq 1, i = 1, \dots, P$ .

**Full/Partial additivity** Full additivity requires the abundances for a mixed pixel to sum to 1, with the implicit assumption that all the endmembers comprising the pixel spectrum in  $\mathbf{x}$  are present in the columns of  $\mathbf{S}$ :  $\sum_{i=1}^P a_i = 1$ . Partial additivity is a generalization that only requires the sum of abundances to be less than or equal to one, and it applies when the set of endmembers in the scene might be incomplete:  $\sum_{i=1}^P a_i \leq 1$ .

Interpretation of data		Non-statistical		Statistical									
Description of Randomness		Non-parametric				Parametric							
Optimization Criterion		Non-squared-error		Squared-error		ML				MAP			
Specifics	Outputs	Continuous-valued abundances	Discretized abundances	Continuous-valued abundances		Discretized abundances				Continuous-valued abundances			
	Inputs	Deterministic endmembers		Endmembers not required		Deterministic endmembers	Stochastic endmembers	Endmembers not required		Deterministic endmembers			
	Noise	Additive obs. noise	No noise model	Additive obs. noise		Additive Gaussian obs. noise	Additive Gaussian obs. noise	No noise model		Additive Gaussian obs. noise			
Algorithm	FVA	Confidence Estimates	ULS	ULS	ULS	ULS	ULS	Projection	NMF	Gaussian-based Estimation	Classical Inverse		
		Confidence Estimates	ULS	ULS	ULS	ULS	ULS	Projection	NMF	Gaussian-based Estimation	Classical Inverse		
Required A Prior Info.	Ground-truth Training Pixels	No	No	No	No	No	No	No	No	No	No	No	No
	Noise Statistics	No	No	No	No	No	No	No	No	No	No	No	No
	Reference Endmember Spectra	No	No	No	No	No	No	No	No	No	No	No	No
Abundance Constraints	Full Additivity	No	Yes	No	Yes	No	No	Yes	No	Yes	No	Yes	Yes
	Non-negativity	No	Yes	No	Yes	No	No	No	Yes	No	Yes	Yes	Yes
	Purity	No	Yes	No	No	No	No	No	Yes	No	Yes	Yes	Yes
Additional Outputs	Priors for Endmembers	No											
	Endmembers	No											
Processing	Iterative	No	Yes	No	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes
	Block-pixel Processing	No	Yes	No	No	No	No	No	No	Yes	Yes	Yes	No
	Requires Inverse	Yes	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes
	Requires SVD/PCA	No											

Figure 5. Taxonomy of inversion algorithms.

### 3.4.1 Non-statistical and Non-parametric Inversion

In this section, we examine inversion algorithms that are neither statistical nor presume a parametric formulation for the data. The abundances are considered to be deterministic.

**3.4.1.1 Non-squared-error** We now examine methods for inversion that do not rely on squared-error, but, instead, explore practical aspects of inversion while addressing some or all of the physical constraints.

**Filter Vector Algorithm** The Filter Vector Algorithm (FVA) [6] is a module of ORASIS that performs inversion by designing an ensemble of filters that are in one-to-one correspondence with each endmember in  $\mathbf{S}$ . Crucial to producing this result is that the columns of  $\mathbf{S}$  must be linearly independent. When the spectrum of a mixed pixel is correlated by one filter, the output is an estimate of the fractional abundance belonging to the corresponding endmember. Thus, if  $\mathbf{F}$  is an  $M \times P$  matrix whose columns are filters, two desirable properties are  $\mathbf{F}^T \mathbf{S} = \mathbf{I}_P$  and  $\mathbf{F}^T \mathbf{j} = \mathbf{0}$ , where  $\mathbf{j}$  is an  $M \times 1$  vector of ones. The first condition requires the filters to be orthogonal to all endmembers except one. The second condition requires that each filter be zero-mean, i.e., the filters should reject all DC terms in the pixel being unmixed and is useful when  $\mathbf{w}$  in (1) has a non-zero mean. The solution to  $\mathbf{F}$  can be shown to be

$$\mathbf{F} = \tilde{\mathbf{S}}((\tilde{\mathbf{S}}^T \mathbf{S})^{-1})^T \quad (19)$$

where  $\mathbf{F}$  requires the mean-removed matrix of endmembers,  $\tilde{\mathbf{S}} = (\mathbf{I}_M - \frac{1}{M} \mathbf{J}^T) \mathbf{S}$ , and  $\mathbf{J}$  is a matrix of ones.

**I-Divergence** As stated earlier, a key to unmixing is finding mathematical structures amenable to the physical limitations inherent to the problem. Mixed pixel spectra are non-negative, and, consequently, a distance measure amenable to comparing non-negative data is the I-divergence which compares two non-negative deterministic functions and yields a non-negative measure of dissimilarity having a value of zero when the functions are identical [28]. For two non-negative vectors of length  $M$ ,  $\mathbf{x}(1)$  and  $\mathbf{x}(2)$ , the I-divergence is defined by

$$I(\mathbf{x}(1), \mathbf{x}(2)) = \sum_{i=1}^M x_i(1) \ln \left[ \frac{x_i(1)}{x_i(2)} \right] - \sum_{i=1}^M (x_i(1) - x_i(2)). \quad (20)$$

Using the block notation for  $N$  pixels and given  $\mathbf{X}$  and  $\mathbf{S}$ , the inversion algorithm selects the elements of  $\mathbf{A}$  for the  $N$  pixel spectra in  $\mathbf{X}$  so that, based on (20), they minimize

$$I(\mathbf{A}) = \sum_{j=1}^N I(\mathbf{x}_j, \sum_{i=1}^M A_{ij} \mathbf{s}_i). \quad (21)$$

Imposed on the columns of  $\mathbf{A}$  during the optimization is the full additivity constraint.  $I(\mathbf{A})$  is minimized by a recursion that generates successive values for the entries of  $\mathbf{A}$ . Although the

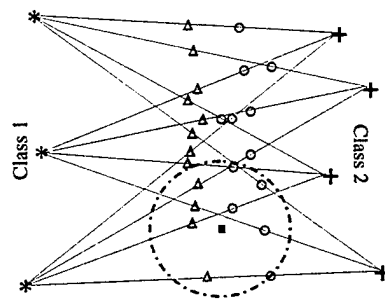


Figure 6. Confidence measurement for 50%-50% mixture (triangles) and 30%-70% (circles) mixture.

algorithm does not incorporate a statistical noise model, each estimate for  $\mathbf{A}$  upholds the non-negativity, purity, and full additivity constraints.

**Confidence Estimates** Most inversion algorithms utilize one endmember to represent the spectrum of a particular substance, or class. Depending on the observation conditions, however, one class may require multiple spectra to satisfactorily capture its variability. The possibility of intra-class variability, and its impact on inversion, is addressed for two-class problems using a non-parametric technique that estimates abundances based on the geometric proximity of spectral observations to the spectrum belonging to each class [25].

This approach is best illustrated graphically. In Figure 6, consider the pixel to be unmixed, denoted by a filled square, using two classes. The various instantiations of each class are denoted by  $A$  asterisks and  $B$  pluses, respectively. The mixed pixel may then be a linear combination of any pair of observations arising from the two classes. A line between a pair indicates the locus of values that a mixed pixel may have, and the location on the line is indicative of the percentage of each component. For instance, the triangles indicate positions on the lines having 50% of each class, whereas circles indicate a 30%-70% relationship. If a circle of appropriate size is drawn around around the pixel to be unmixed, confidence estimates may be derived for each possible mixing proportion by counting the number of occurrences falling within the circle and dividing by the total number of possible model mixtures,  $AB$ . Based on these ratios, a pixel is assigned the abundances yielding the highest confidence estimate.

Abundance estimates based on this technique automatically obey the non-negativity, purity, and full additivity constraints. Since an infinite number of confidence estimates can be generated, a discretized lattice of abundance estimates most likely would be used to obtain confidence estimates.

**3.4.1.2 Squared-error** We now investigate solutions that are based on the method of least squares. Building upon the unconstrained least squares estimate, inversion algorithms striving to minimize squared-error address physical constraints detailed earlier as well as realistic limitations

that arise from practical scenarios. Computational approaches for minimizing quadratic cost functions with linear constraints fall into the realm of quadratic programming and are not discussed here.

**Unconstrained Least Squares** The theory of least squares yields an estimate,  $\hat{\mathbf{a}}$ , that minimizes the total squared-error between  $\mathbf{x}$  and  $\mathbf{S}\hat{\mathbf{a}}$ ,  $\hat{\mathbf{a}} = \min_{\mathbf{a}} |\mathbf{x} - \mathbf{S}\mathbf{a}|^2$ . The unconstrained least squares estimate,  $\hat{\mathbf{a}}^U$ , is given by [31]

$$\hat{\mathbf{a}}^U = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{x}. \quad (22)$$

Implicit in this estimate are two assumptions. First,  $M > P$ , which is typical in hyperspectral systems. Second,  $\mathbf{S}$  is an  $M \times P$  matrix with full column rank so that  $(\mathbf{S}^T \mathbf{S})^{-1}$  exists.

**Full Additivity** The addition of the full additivity constraint to the least squares problem restricts the optimal estimate of  $\hat{\mathbf{a}}$  to reside on a hyperplane (a surface whose dimensionality is one less than the signal dimension). This is stated formally as

$$\hat{\mathbf{a}}^F = \min_{\mathbf{a}} |\mathbf{x} - \mathbf{S}\mathbf{a}|^2, \quad \sum_{i=1}^P a_i = 1. \quad (23)$$

The general least squares estimate with linear constraints is derived using Lagrange multipliers to be [16]

$$\hat{\mathbf{a}}^F = \hat{\mathbf{a}}^U - (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{Z}^T [\mathbf{Z} (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{Z}^T]^{-1} (\mathbf{Z} \hat{\mathbf{a}}^U - \mathbf{b}). \quad (24)$$

$\hat{\mathbf{a}}^F$  is the least squares estimate of the abundances when the full additivity constraint is enforced,  $\mathbf{Z}$  is an  $R \times P$  matrix and  $\mathbf{b}$  is an  $R \times 1$  vector that together convey the  $R$  linearly independent equations,  $\mathbf{Z}\mathbf{a}^F = \mathbf{b}$ , that serve as linear constraints on the solution of (1). To enforce full additivity,  $\mathbf{Z}$  is a  $1 \times P$  row vector whose entries are all ones, and  $\mathbf{b} = 1$ . Examining (24) reveals that the solution enforcing full additivity consists of the unconstrained LSE solution,  $\hat{\mathbf{a}}^U$ , with an additive correction term that depends on the matrix of endmembers,  $\mathbf{S}$ , and the error incurred by the unconstrained solution,  $\hat{\mathbf{a}}^U$ , in satisfying the full additivity constraint.

**Projection on Convex Sets** One problem with the unconstrained least squares estimate in (22) happens when  $\mathbf{S}$  does not possess full column rank. In hyperspectral sensing, this occurs when the endmembers are linearly dependent, a situation that is encountered when pixels are inverted using spectrally similar endmembers.

This problem is addressed by defining two convex sets, one based on  $\mathbf{S}$  and the other by the non-negativity constraint, and by recognizing that any  $\mathbf{a}$  which solves (1) must necessarily obey two properties [34]. First, the entries of  $\mathbf{a}$  must be non-negative to be physically meaningful. Second, any valid  $\mathbf{a}$  should be expressible as a sum of two components,  $\hat{\mathbf{a}} = \mathbf{a}^c + \mathbf{a}^n$ . The first term,  $\mathbf{a}^c$

resides in the column space of  $\mathbf{S}$ , and  $\mathbf{a}^n$  resides in its null space. The non-negativity constraint on  $\hat{\mathbf{a}}$ , can be imposed by adding a requirement on  $\mathbf{a}^n$ .

$$a_i^n \geq -a_i^c, \quad 1 \leq i \leq P. \quad (25)$$

The objective is to find a value for  $\mathbf{a}^n$  that resides in the null space of  $\mathbf{S}$  – this assures that  $\mathbf{S}\mathbf{a}^n$  contributes nothing to the mixed pixel spectrum – but also resides in the convex space of all possible vectors having entries greater than the negative of the corresponding entries in  $\mathbf{a}^c$ , assuring the non-negativity of the entries in  $\hat{\mathbf{a}}$ .

An iterative algorithm determines  $\mathbf{a}^n$  so that it resides in the intersection of the convex set defined by the null space of  $\mathbf{S}$  and the convex set of values satisfying (25). The algorithm alternately projects solutions for  $\mathbf{a}$  derived from one convex set onto the other convex set until a standard stopping criterion is met. Successive values for  $\mathbf{a}^n$  converge in the intersection of the two convex sets.

### 3.4.2 Statistical and Non-parametric Inversion

In this section we consider inversion algorithms that utilize statistics to invert pixels but do not rely on parametric expressions to model their random behavior.

**3.4.2.1 Squared-error** We examine several inversion algorithms in this section that minimize squared-error.

**Minimum Variance Unbiased Estimator** If  $\mathbf{w}$  is a zero-mean random process in (1) and has a covariance,  $\Gamma_w$ , then the minimum variance estimate,  $\hat{\mathbf{a}}^V$ , is defined as

$$\hat{\mathbf{a}}^V = (\mathbf{S}^T \Gamma_w^{-1} \mathbf{S})^{-1} \mathbf{S}^T \Gamma_w^{-1} \mathbf{x}. \quad (26)$$

This solution is based on the weighted least squares approach [16] and resembles  $\hat{\mathbf{a}}^U$  in (22), with the common caveat that  $\mathbf{S}$  must have full column rank. Incorporating second-order statistics of the random noise,  $\mathbf{w}$ ,  $\hat{\mathbf{a}}^V$  minimizes  $(\mathbf{x} - \mathbf{S}\hat{\mathbf{a}})^T \Gamma_w^{-1} (\mathbf{x} - \mathbf{S}\hat{\mathbf{a}})$  rather than  $(\mathbf{x} - \mathbf{S}\hat{\mathbf{a}})^T (\mathbf{x} - \mathbf{S}\hat{\mathbf{a}})$ . Because the unbiased estimator in (26) is the best linear unbiased estimator (BLUE) it is called the minimum variance unbiased estimator (MVUE).

**Ground-truth-based Estimators** Ground truth consists of a set of pixel spectra from a scene for which the fractional abundances are known, or estimates exist, and an estimator designed for optimality on this training set may then be used to invert other mixed pixels in the scene with the advantage of deriving estimates without knowledge of the actual endmember spectra in the scene. Two estimators based on this concept, each optimizing a different quantity in the ground truth region, have been proposed to estimate abundances [27].

Before comparing the two techniques, we introduce nomenclature that is useful to both. Adopting block notation for  $N$  training pixels, let the estimate for the abundance correlation matrix be given by  $\hat{\mathbf{R}} = \frac{1}{N} \mathbf{A} \mathbf{A}^T$ , let the estimate of the endmembers be given by  $\hat{\mathbf{S}} = (\frac{1}{N} \mathbf{X} \mathbf{A}^T) \hat{\mathbf{R}}^{-1}$ ,

and let the noise covariance estimate be defined as  $\hat{\Gamma}_w = \frac{1}{N}(\mathbf{X} - \hat{\mathbf{S}}\mathbf{A})(\mathbf{X} - \hat{\mathbf{S}}\mathbf{A})^T = \frac{1}{N}\mathbf{X}\mathbf{X}^T - (\frac{1}{N}\mathbf{X}\mathbf{A}^T)\hat{\mathbf{R}}^{-1}(\frac{1}{N}\mathbf{A}\mathbf{X}^T)$ . We can also denote the average pixel spectra by  $\bar{\mathbf{x}}$  and the average abundance in the training set by  $\bar{\mathbf{a}}$ . Also, let  $\mathbf{j}$  be a  $P \times 1$  vector of ones, and let  $\mathbf{J} = \mathbf{j}\mathbf{j}^T$ .

The **classical estimator** postulates a physical model (the linear mixture model) for the synthesis of a mixed pixel, estimates the parameters of this model from the training set, then inverts the model to estimate abundances for the remaining image pixels. It minimizes

$$\text{trace}[(\mathbf{X} - \hat{\mathbf{S}}\hat{\mathbf{A}}^C)^T \Gamma_w^{-1}(\mathbf{X} - \hat{\mathbf{S}}\hat{\mathbf{A}}^C)] \quad (27)$$

which is the sum, over all  $N$  training pixels, of the weighted squared-error between  $\mathbf{X}$  and  $\hat{\mathbf{S}}\hat{\mathbf{A}}^C$ . The classical estimator,  $\hat{\mathbf{a}}^C$ , is given by

$$\hat{\mathbf{a}}^C = \bar{\mathbf{a}} + \mathbf{E}^C \hat{\mathbf{S}}^T \hat{\Gamma}_w^{-1}(\mathbf{x} - \bar{\mathbf{x}}). \quad (28)$$

Note that while  $\bar{\mathbf{a}}$  is the average fractional abundance in the training set, it is constrained to uphold the full additivity constraint, i.e., the sum of the entries in  $\bar{\mathbf{a}}$  equals one.  $\mathbf{E}^C$  is the prediction error matrix for the classical estimator and is defined as  $\mathbf{E}^C = \mathbf{U} - \alpha \mathbf{U} \mathbf{J} \mathbf{U}$ , where  $\mathbf{U} = (\hat{\mathbf{S}}^T \hat{\Gamma}_w^{-1} \hat{\mathbf{S}})^{-1}$  and  $\alpha = (\mathbf{j}^T \mathbf{U} \mathbf{j})^{-1}$ .

The **inverse estimator** generates an estimate for  $\mathbf{a}$  that exemplifies the precepts of multivariate regression. Whereas the classical estimator sought to minimize the weighted squared-error between  $\mathbf{X}$  and  $\hat{\mathbf{S}}\hat{\mathbf{A}}^C$ , the inverse estimator minimizes

$$\text{trace}[(\hat{\mathbf{A}}^I - \mathbf{A})^T (\hat{\mathbf{A}}^I - \mathbf{A})] \quad (29)$$

which is the squared-error between  $\mathbf{A}$  and  $\hat{\mathbf{A}}^I$ . The solution for  $\hat{\mathbf{a}}^I$  is given by

$$\hat{\mathbf{a}}^I = \bar{\mathbf{a}} + \mathbf{E}^I \hat{\mathbf{S}}^T \hat{\Gamma}_w^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \quad (30)$$

where, if  $\mathbf{C} = \mathbf{Q} - \mathbf{Q} \mathbf{J} \mathbf{Q}$ ,  $\mathbf{E}^I = \mathbf{C}(\mathbf{C} + \mathbf{U})^{-1} \mathbf{U}$ . Reviewing (28) and (30) reveals that both estimators only differ by their respective prediction error matrices,  $\mathbf{E}^C$  and  $\mathbf{E}^I$ , whose the diagonal elements provide the mean squared prediction errors for the associated ground class type.

### 3.4.3 Statistical and Parametric Inversion

In this section we consider inversion algorithms that utilize statistical, parametric representations of the received data to invert mixed pixels. The principle distinction between algorithms in the section arises from the assumption of randomness for abundances. ML techniques presume the abundances are deterministic quantities, while MAP techniques assume they are random, necessitating a prior density for the abundances.

**3.4.3.1 Maximum Likelihood** We consider inversion algorithms in this section that assume the received spectra obey a parametric model but have deterministic abundances. In all cases, each algorithm maximizes a forward density,  $p_{\mathbf{x}|\mathbf{a}}(\mathbf{x}|\mathbf{a})$ , to obtain an ML estimate for the  $\mathbf{a}$ .



**Gaussian Minimum Variance Unbiased Estimate** The performance of the MVUE in Section 3.4.2.1 can be further quantified if the noise,  $\mathbf{w}$ , in (1) is zero-mean and also Gaussian. In this case, the estimator,  $\hat{\mathbf{a}}^{\text{GV}}$ , equals  $\hat{\mathbf{a}}^{\text{V}}$  in (26). In addition to the previously enumerated properties,  $\hat{\mathbf{a}}^{\text{GV}}$  also achieves the Cramer-Rao lower bound for unbiased estimators.

**Weighted Mahalanobis Distance** When the additive noise,  $\mathbf{w}$ , in (1) is Gaussian, an ML estimator finds  $\mathbf{a}$  that maximizes a Gaussian forward density,  $p_{\mathbf{x}|\mathbf{a}}(\mathbf{x}|\mathbf{a})$ . This is equivalent to minimizing the Mahalanobis distance, which is defined as  $(\mathbf{x} - \mathbf{S}\mathbf{a})^T \Gamma_{\mathbf{w}}^{-1} (\mathbf{x} - \mathbf{S}\mathbf{a})$ .

On the other hand, the uncertainty in  $\mathbf{x}$  introduced by  $\mathbf{w}$  may be alternately interpreted as the uncertainty in the endmember spectra in  $\mathbf{S}$  [18]. This alternative to the signal model in (1) is expressed as  $\mathbf{x} = (\mathbf{S} + \mathbf{E})\mathbf{a}$ , where  $\mathbf{E}\mathbf{F}$  is an  $M \times P$  matrix whose columns,  $\mathbf{e}_i, i = 1, \dots, P$ , are  $M$ -dimensional Gaussian vectors having mean,  $\mu_{s_i}$ , and covariance,  $\Gamma_{s_i}$ . The covariance for  $p_{\mathbf{x}|\mathbf{a}}(\mathbf{x}|\mathbf{a})$ ,  $\Gamma_{\mathbf{x}|\mathbf{a}}$ , is now a sum of covariances and cross-covariances, each weighted by the associated fractional abundances,  $\Gamma_{\mathbf{x}|\mathbf{a}} = \sum_{i=1}^P \sum_{j=1}^P a_i a_j \Gamma_{s_i s_j}$ . The dependence of  $\Gamma_{\mathbf{x}|\mathbf{a}}$  on  $\mathbf{a}$  makes an ML optimization with respect to  $\mathbf{a}$  considerably more complicated than the standard minimization of the Mahalanobis distance. The log-likelihood function to be minimized,  $\Lambda(\mathbf{x})$ , derived from  $p_{\mathbf{x}|\mathbf{a}}(\mathbf{x}|\mathbf{a})$ , is

$$\Lambda(\mathbf{x}) = \ln |\Gamma_{\mathbf{x}|\mathbf{a}}| + (\mathbf{x} - \mathbf{S}\mathbf{a})^T \Gamma_{\mathbf{x}|\mathbf{a}}^{-1} (\mathbf{x} - \mathbf{S}\mathbf{a}). \quad (31)$$

Invoking the full additivity constraint to reduce the number of variables by one, the minimization of (31) involves finding a  $(P - 1)$ -dimensional subspace in  $\Re^M$  that contains  $\mathbf{x}$  as well as a projection of  $\mathbf{s}_k, k = 1, \dots, P$ . The goal is to iteratively minimize the metric in (31) by parameterizing all possible  $P - 1$  subspaces and finding the best family of basis vectors. The method, however, does not inherently enforce the non-negativity and purity constraints.

**3.4.3.2 Maximum A Posteriori** In this section, we depart from the assumption that the abundances are deterministic quantities. Several ML algorithms resulted in optimal solutions when  $\mathbf{w}$  is Gaussian and  $\mathbf{a}$  is assumed to be deterministic. If, however, the abundances are now considered to be random, a maximum a posteriori (MAP) formulation may be employed to unmix pixels. The general form of this estimator is denoted by

$$\max_{\mathbf{a}} p_{\mathbf{a}|\mathbf{x}}(\mathbf{a}|\mathbf{x}) = \max_{\mathbf{a}} p_{\mathbf{x}|\mathbf{a}}(\mathbf{x}|\mathbf{a}) \frac{p_{\mathbf{a}}(\mathbf{a})}{p_{\mathbf{x}}(\mathbf{x})}. \quad (32)$$

Invoking a standard assumption, we can ignore  $p_{\mathbf{x}}(\mathbf{x})$  since it has no dependence on  $\mathbf{a}$ . The construction of  $p_{\mathbf{a}}(\mathbf{a})$ , however, is critical in the following algorithms, as well as in Bayesian inference in general.

**Fuzzy Membership** We consider an algorithm for inverting mixed pixels that is a generalization of a classification technique that permits a pixel to have fractional membership in multiple classes [17]. Appropriately constrained, the fractional memberships are construed as abundance

coefficients, and the Bayesian model in (32) is employed to estimate  $\mathbf{a}$ . The approach assumes that the abundance vector,  $\mathbf{a}$ , is random and Gaussian, and the columns of  $\mathbf{S}$  are the average spectra for  $P$  distinct, known classes.

For a system of  $P$  abundances, the dimensionality of the solution space for  $\mathbf{a}$  is reduced from  $P$  to  $(P - 1)$  by imposing the full additivity constraint. Accordingly,  $\mathbf{a}$ , is transformed to a new  $(M - 1) \times 1$  vector,  $\mathbf{v}$ , through one of two possible differencing transformations. The resulting distributions for both evaluations of  $\mathbf{v}$  are approximated by the quasi-intrinsic Gaussian conditional autoregression (QICAR) which asserts that the density of  $\mathbf{v}$  is equivalent to the difference between Gaussian-distributed membership vectors in a symmetric neighborhood around the pixel in question. In essence, the differenced abundance vector,  $\mathbf{v}$ , for a single pixel is approximated by differences in class membership for a surrounding neighborhood of pixels. Then, since conditioning on  $\mathbf{a}$  is equivalent to conditioning on  $\mathbf{v}$ , the Bayesian formulation in (32) can be optimized with  $\mathbf{v}$  in place of  $\mathbf{a}$ .

The derivative of (32) with respect to  $\mathbf{v}$  is performed using the Gaussian QICAR distributions for the prior. The resulting expressions can be reworked to yield an iterative structure which successively estimates  $\mathbf{v}$ , from which  $\mathbf{a}$  can be obtained. However, the enforcement of the full additivity constraint does not preclude values of  $\mathbf{a}$  from being negative or greater than one. This is corrected by projecting  $\mathbf{a}$  to the nearest point in the space of acceptable values.

**Log-Odds Transformation** Enforcing the non-negativity and purity constraints requires a constrained optimization of (32). This problem is addressed by returning the constrained optimization to an unconstrained one by warping the parameter space so that techniques intended for unconstrained optimizations may be used [29]. The approach assumes  $\mathbf{w}$  in (1) has a multivariate Gaussian distribution. Acknowledging uncertainty in the exact values in (1), the full additivity constraint is relaxed, and  $p_{\mathbf{a}}(\mathbf{a})$  is modeled as a univariate Gaussian distribution with a small variance,  $\sigma_{\mathbf{a}}^2$ , describing the sum of the abundances as

$$p_{\mathbf{a}}(\mathbf{a}) \propto \exp\left(-\frac{(h(\mathbf{a}) - 1)^2}{\sigma_{\mathbf{a}}^2}\right) \quad (33)$$

where  $h(\mathbf{a}) = \sum_{i=1}^P a_i$ . Note that  $a_k$  ranges from  $(-\infty, \infty)$ , which contradicts the non-negativity and purity constraints on the abundances. This will be addressed shortly.

The posterior density can then be expressed as

$$p_{\mathbf{a}|\mathbf{x}}(\mathbf{a}|\mathbf{x}) \propto \exp\left(-\frac{(h(\mathbf{a}) - 1)^2}{\sigma_{\mathbf{a}}^2} - (\mathbf{x} - \mathbf{S}\mathbf{a})^T \Sigma_w^{-1} (\mathbf{x} - \mathbf{S}\mathbf{a})\right) \quad (34)$$

where the optimal solution for  $\mathbf{a}$  maximizes this posterior density. However, by (34), the optimal value for  $\mathbf{a}$  may reside outside of  $[0, 1]$ . This possibility was introduced by Gaussian densities having non-zero probabilities for all values of  $\mathbb{R}$ . Instead, a desirable operation would apply a reversible transformation to  $\mathbf{a}$ , transforming the interval,  $[0, 1]$ , to  $[-\infty, \infty]$ . Such a warping of the parameter space would return the problem to an unconstrained optimization in the transformed space.

Let the transformation defining  $a_k^*$  be defined as

$$a_k^* = \log\left(\frac{a_k}{1 - a_k}\right), \quad k = 1, \dots, P. \quad (35)$$

If  $\mathbf{a}^*$  admits a Gaussian distribution, then the posterior density for  $\mathbf{a}^*$  can be written as

$$p_{\mathbf{a}^*|\mathbf{x}}(\mathbf{a}^*|\mathbf{x}) \propto \exp\left(-\frac{(h^*(\mathbf{a}^*) - 1)^2}{\sigma_{a^*}^2} - (\mathbf{x} - \mathbf{S}^*\mathbf{a}^*)^T \Sigma_w^{-1} (\mathbf{x} - \mathbf{S}^*\mathbf{a}^*)\right). \quad (36)$$

Both  $h^*(\cdot)$  and  $\mathbf{S}^*$  are operators that incorporate the inverse transformation of (35) as well as  $h(\cdot)$  and  $S$  respectively. The prior distribution, although its arguments have not been transformed by (35) is retained since it is also valid over  $(-\infty, \infty)$ . Newton's method is discussed as a means of maximizing (36).

**Maximum Entropy** The Bayesian formulation in (32) may be reconsidered from the viewpoint of entropy [7]. Probability distributions are required to be non-negative and sum to one, a property shared by the full additivity constraint. If  $\mathbf{a}$  is a  $P \times 1$  vector of fractional abundances belonging to a pixel, the corresponding entropy,  $H(\mathbf{a})$ , is

$$H(\mathbf{a}) = - \sum_{i=1}^P a_i \ln a_i. \quad (37)$$

The entropy derived from a vector of abundances,  $\mathbf{a}$ , for a mixed pixel can be shown to be, through a combinatoric approximation, proportional to the number of ways a pixel may contain endmembers in the proportions denoted by  $\mathbf{a}$ . Combining this fact with another approximation, it can be shown that  $p_{\mathbf{a}}(\mathbf{a})$  is proportional to  $H(\mathbf{a})$ .

Hence, for a particular mixed pixel, the objectives are to find the abundances that maximize (37) and to satisfy the obvious requirement that  $\mathbf{x} = \mathbf{S}\mathbf{a}$  in the presence of additive noise. Taken separately, these goals translate to expressions for prior and forward densities, respectively, which may be multiplied to yield the posterior density in (32). The log-posterior density is

$$\ln p_{\mathbf{a}|\mathbf{x}}(\mathbf{a}|\mathbf{x}) \propto H(\mathbf{a}) - \frac{K}{2} (\mathbf{x} - \mathbf{S}\mathbf{a})^T \Sigma_w^{-1} (\mathbf{x} - \mathbf{S}\mathbf{a}). \quad (38)$$

$K$  is a constant that is included to account for the approximation linking  $p_{\mathbf{a}}(\mathbf{a})$  to  $H(\mathbf{a})$ . Thus, the overall objective function to be optimized is a sum of an entropy calculation and a Mahalanobis distance.

#### 4. ALGORITHM TAXONOMY: NON-LINEAR UNMIXING

We have not presented a taxonomy for non-linear unmixing algorithms for two reasons. First, the high-level categories in Figure 2 are predicated on the the linear mixing model in (1) and the properties induced by its amenability to practical mathematical and statistical tools. Approaches to non-linear unmixing are driven by the physically rigorous models in Section 2.2, which do not easily accept the tools apparent in linear unmixing. The structure in Figure 2 is primarily predicated on mathematical and statistical considerations. In this sense, there is a mutual orthogonality that inhibits physics-based analysis from co-existing with simple, structured mathematical algorithms.

Second, and most importantly, in comparison to the number of algorithms aimed at linear unmixing, very little research has been directed toward tackling the non-linear problem in a systematic way. Some of the efforts are still exploring the sufficiency of the linear paradigm and have yielded useful experimental results [21, 24]. Others have meticulously developed physical mixing models from the perspectives outlined in Section 2.2; in these cases, their unwieldiness is mitigated by common-sense physical approximations that permit simple mathematical techniques to estimate fractional abundances [8, 14, 21]. For example, (2) can be solved for  $\mathbf{a}$  using the method of least squares, which is documented as an inversion technique in Figure 4. As non-linear unmixing techniques mature, a taxonomy comparable to the ones in Figures 3-5, but built on physical principles, will naturally emerge.

## 5. PERFORMANCE COMPARISONS

Spectral unmixing has emerged as a key application arising from the wealth of spectral measurements in hyperspectral processing. Several communities have shown great interest in the compositional analysis of mixed pixels. Unmixing attempts to decompose mixed pixels in terms of distinct, unique substances, and provide a foundation for doing sub-pixel material identification.

We undertake this comparison of unmixing algorithm performance with the knowledge that many algorithms exist, and new methods are constantly being explored and tested. Several disciplines are participating in the attempt to perform unmixing, such as geology, geophysics, engineering, and analytical chemistry.

### 5.1 DATA

For the purpose of our comparisons, we utilize a natural scene that was imaged by the HYDICE sensor. The scene in Figure 7 was collected at approximately noon on September 22, 1997 as part of the Alpine Radiance I data collection in Northern California and consists of 320 lines of data, each having 306 samples with 210 spectral bands. The scene is a gently sloping open area at an elevation of 8500 feet that is bordered by both mature and young lodgepole pine trees, as well as areas of granite-derived loams that are covered in places by low-lying sage brush, aspen trees, and cedar trees. There were no clouds during the data collection and visibility was deemed to be high.

Of the 210 spectral bands, 63 bands were discarded because they reside in the electromagnetic intervals where water vapor absorption is dominant, and the subsequent signal-to-noise ratio (SNR) is too low. Plots of spectra will have gaps to represent the absence of data in those intervals. The radiance measurements for the remaining 147 bands were also corrected for a problem with the absolute calibration of the in-flight calibration bulb that affected the measurements in the 1700 - 2500 nm range. To rectify this, a coefficient vector was applied to the radiance data before atmospheric compensation was performed on the scene. Atmospheric compensation was performed on the scene in Figure 7 to recover surface reflectance values.



*Figure 7. Alpine Radiance Scene I. The data was collected around noon on September 22, 1997.*

## 5.2 SELECTED ALGORITHMS

In this comparison of algorithm performance, we restrict the set of algorithms that we consider to only those that are representative of a fundamental approach, and ignore approaches that are variations on the same concept. Moreover, although it is the first stage of unmixing, we do not compare any algorithms that perform dimension reduction, but instead present some basic statistical information about the scene in Figure 7 that provides insight on the algorithm performance for endmember determination and inversion.

## 5.3 APPROACH

Unfortunately, the ground truth that exists for the scene in Figure 7 is insufficient to corroborate the results from unmixing algorithms. The lack of accurate sub-pixel measurements is a significant problem in trying to assess the absolute performance of unmixing algorithms. In spite of this, we attempt to document the behavior of common algorithms and make judgments about their ability to achieve goals that are desirable for *any* unmixing algorithm.

## 5.4 SCENE STATISTICS

To provide a statistical foundation for the scene in Figure 7, we can first investigate its statistical properties. The scene is comprised of 97920 pixels, characterized by 147 reflectance samples. In Figure 8, the covariance, the cumulative, normalized eigenvalues, and the first five eigenvectors are plotted.

It is clear from Figure 8(b) that the effective dimensionality of the spectra in the scene is relatively low. In order to account for 99% of the power in the scene, only three principal components are necessary. For reference, the first five eigenvectors for the scene are also plotted in Figure 8(c).

## 5.5 ENDMEMBER DETERMINATION

We will compare two techniques for autonomously determining the endmembers in a scene. The first is a technique based on geometric interpretation of the data in the scene, and exploits the parallelism between the linear mixing model (LMM) and the theory of convex sets [3, 5, 9]. The second is based on the idea of clustering the data and finding abundances and endmembers that minimize an overall cost function.

### 5.5.1 Geometric Methods

In our experiment, our goal was to determine the endmembers from the vertices of a simplex that is shrinkwrapped around the high-dimensional volume defined by the data in the scene. Geometric methods identify endmembers in a scene from the vertices of a minimum-volume simplex

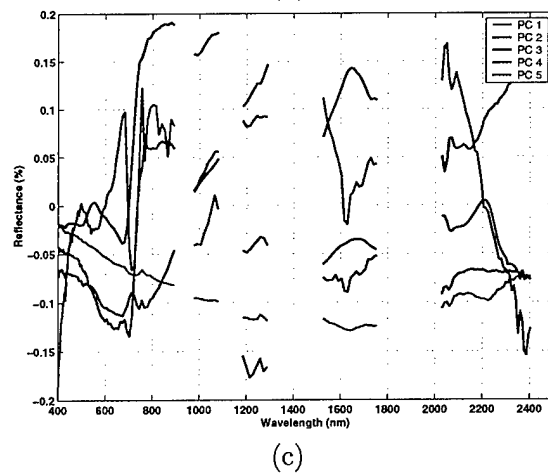
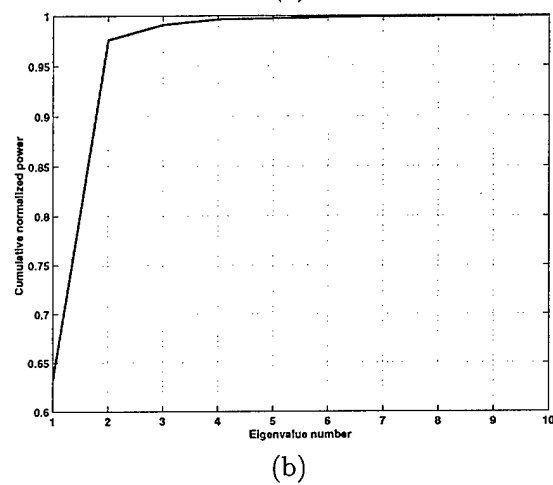
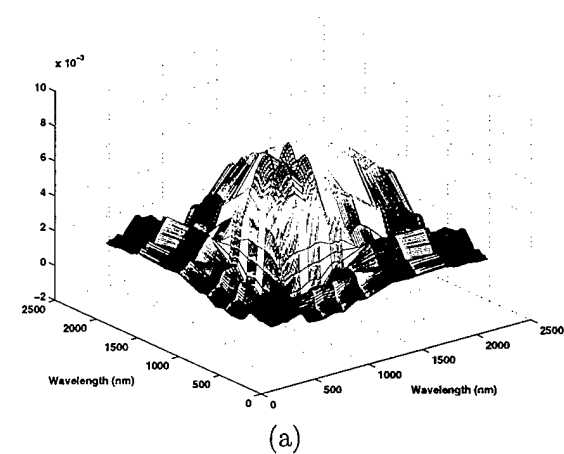


Figure 8. (a) Covariance of scene; (b) Cumulative, normalized eigenvalues; (c) First five principal components



that encapsulates the data. A simplex must necessarily have one more vertex (and facet) than the dimension of the data it encloses, and thus, the number of endmembers is one greater than the effective dimension of the data.

Using the results in Figure 8, we tested the geometric method under the different assumptions that the effective dimension of the data was 3, 4, 5, and 6. Based on the results in Section 5.4, we initially presumed that the approximate dimension of the data was three, and we reduced the original 147-dimensional data to three dimensions using a linear transform constructed from the first three principal components of the covariance. Numerous variations exist on the algorithm, especially in the dimension reduction performed on the data, we felt it was reasonable to use PCA, since it is optimal when optimally preserving the energy in the signals [35].

After reducing the dimension to three, we supplied the data to a convex hull algorithm. We used the QHULL algorithm (<http://www.geom.umn.edu/software/qhull/>) to implement this procedure and then submitted the convex hull of the 3-dimensional points in the scene to a shrinkwrapping algorithm based on the formulation of the Fixed-Point Free Transform (FPFT) [9]. The four endmember reflectance spectra produced by this technique are depicted in Figure 9.

The increasing number of endmembers in each plot from Figure 9 highlights a new, distinct spectra from the scene in Figure 7. As the number increases, new classes appear that were originally lumped in with other classes for lower-dimensional trials of the algorithm.

### 5.5.2 Clustering

In contrast to the geometric technique in Section 3.3.1.1 for endmember determination, clustering techniques have employed iterative, statistical methods to arrive at estimates of endmembers as well as abundances. Despite the variations, the large majority of approaches arrive at estimates using second-order statistics to define statistical classes for each endmember and by relating the fractional abundances of each pixel to partial membership in each class. Then, upon iteration, estimates of abundances and endmembers minimize a cost function. When the cost function arises from a probabilistic, Gaussian formulation, solutions often provide Maximum Likelihood (ML) solutions [17, 30, 33].

We consider for our experiments a non-probabilistic, statistical technique that provides the foundation for many of the clustering-based approaches to unmixing. The idea of fuzzy K-means clustering [10] attempts to optimize a least-squares-based cost function by finding the best endmembers and abundances. For a collection of  $N$  pixels, the objective function may be expressed as

$$J_m(\mathbf{A}, \mathbf{S}) = \sum_{i=1}^P \sum_{j=1}^N (A_{ij})^m (d_{ij})^2 \quad (39)$$

where  $A_{ij}$  is an estimate of the  $i$ -th abundance for the  $j$ -th pixel and  $d_{ij}$ , the weighted squared-error between the  $j$ -th pixel and the  $i$ -th centroid, or endmember,  $\mathbf{s}_i$ , is

$$(d_{ij})^2 = (\mathbf{x}_j - \mathbf{s}_i)^T \mathbf{W} (\mathbf{x}_j - \mathbf{s}_i). \quad (40)$$

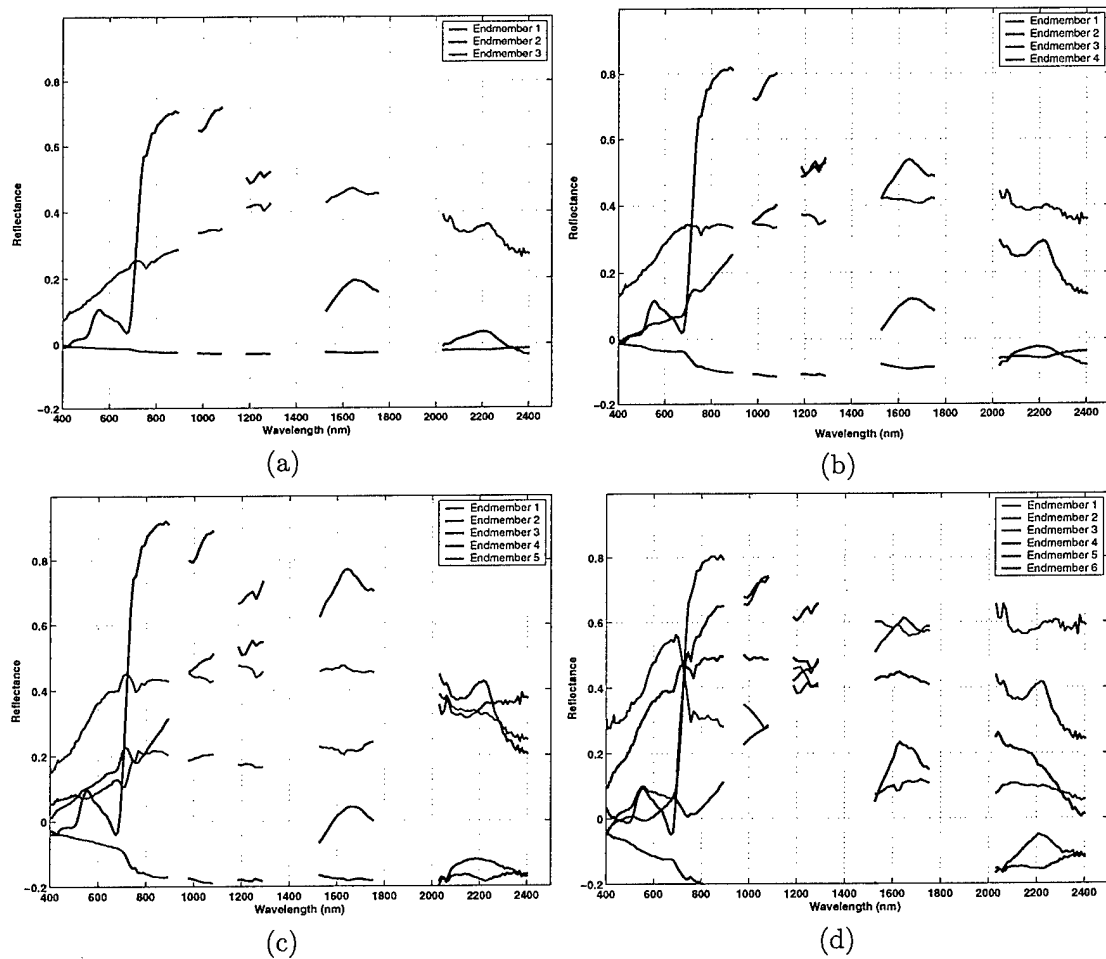


Figure 9. Endmembers for geometric technique using (a) 3-cornered simplex; (b) 4-cornered simplex; (c) 5-cornered simplex; (d) 6-cornered simplex.

In our experiments, we let  $W = I$ , and, upon the recommendation of previous investigators, let  $r = 2.5$ . The endmembers that result for the initial assertion that there are three classes appear in Figure 10.

## 5.6 INVERSION

Endmember determination techniques identify the unique spectra that define the components of the LMM. The complementary task of unmixing is to identify the proportions in which the endmembers appear for each pixel in the scene being analyzed. Numerous algorithms attempt to use pre-determined endmembers to estimate abundances, but the vast majority of techniques are based on the method of least squares, which attempts to minimize  $\|\mathbf{x} - \mathbf{S}\hat{\mathbf{a}}\|^2$ .

A vital part of inversion is including two important, physical constraints. The full additivity constraint requires that  $\sum_{i=1}^P a_i = 1$ , and the non-negativity constraint requires that  $a_i \geq 0, i = 1, \dots, P$ . Several efforts have been made to incorporate these constraints into the unconstrained least squares solution,  $\hat{\mathbf{a}}^U = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{x}$  [31]. In addition to the unconstrained least squares solution, we will consider two more inversion methods in our performance comparisons. The full additivity least squares solution [16] is given by

$$\hat{\mathbf{a}}^F = \hat{\mathbf{a}}^U - (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{Z}^T [\mathbf{Z} (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{Z}^T]^{-1} (\mathbf{Z} \hat{\mathbf{a}}^U - \mathbf{b}). \quad (41)$$

$\hat{\mathbf{a}}^F$  is the least squares estimate of the abundances when the full additivity constraint is enforced,  $\mathbf{Z}$  is an  $R \times P$  matrix and  $\mathbf{b}$  is an  $R \times 1$  vector that together convey the  $R$  linearly independent equations,  $\mathbf{Z} \mathbf{a}^F = \mathbf{b}$ , that serve as linear constraints on the solution of (1). To enforce full additivity,  $\mathbf{Z}$  is a  $1 \times P$  row vector whose entries are all ones, and  $\mathbf{b} = 1$ . Examining (41) reveals that the solution enforcing full additivity consists of the unconstrained LSE solution,  $\hat{\mathbf{a}}^U$ , with an additive correction term that depends on the matrix of endmembers,  $\mathbf{S}$ , and the error incurred by the unconstrained solution,  $\hat{\mathbf{a}}^U$ , in satisfying the full additivity constraint.

Figure 11 conveys the results of unconstrained least squares unmixing when the six endmembers derived from the geometric endmember algorithm are used (one is discarded for being mostly negative). Also included is a figure illustrating the sum of the estimated abundances. Clearly, the unconstrained least squares version does not sum to one in the case of a majority of pixels in the scene.

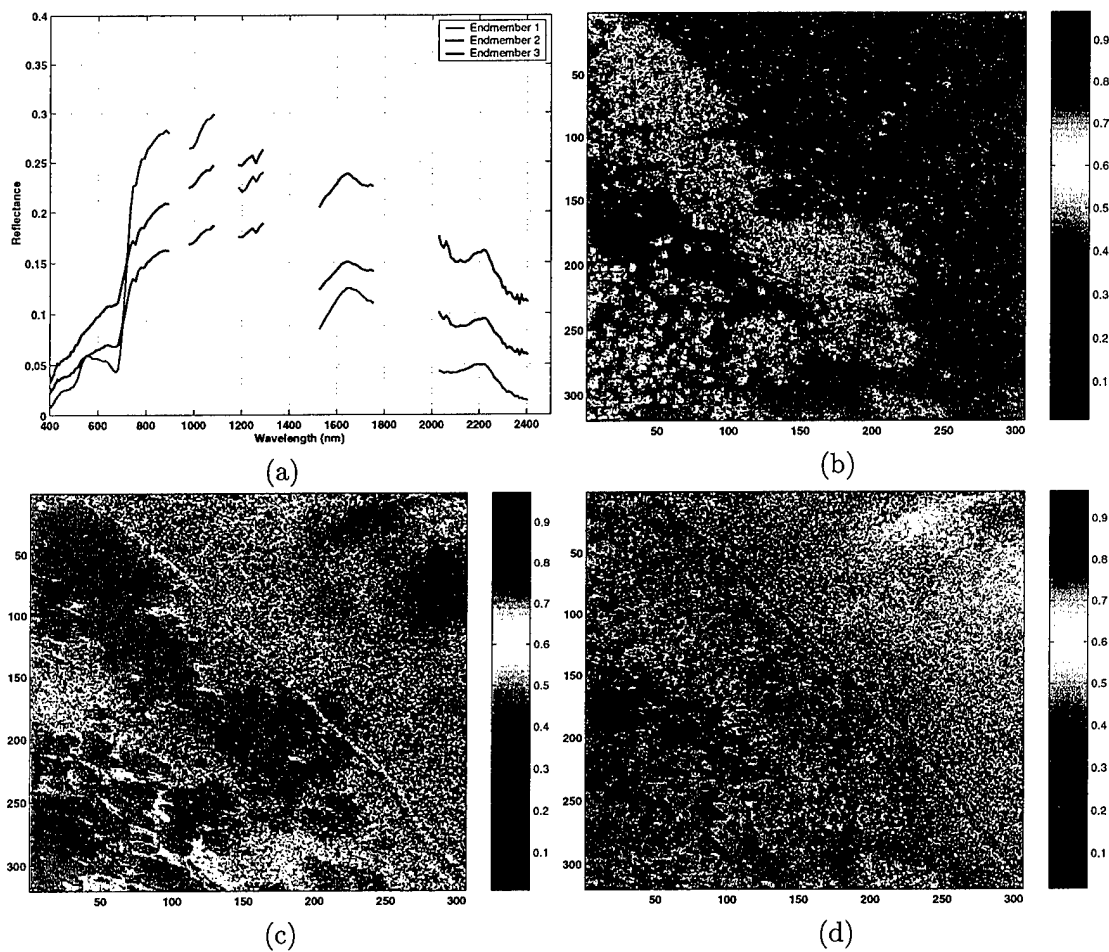


Figure 10. (a) Endmembers for fuzzy clustering algorithm with 3 classes; (b) abundance plane for Endmember 1; (c) abundance plane for Endmember 2; (d) abundance plane for Endmember 3.

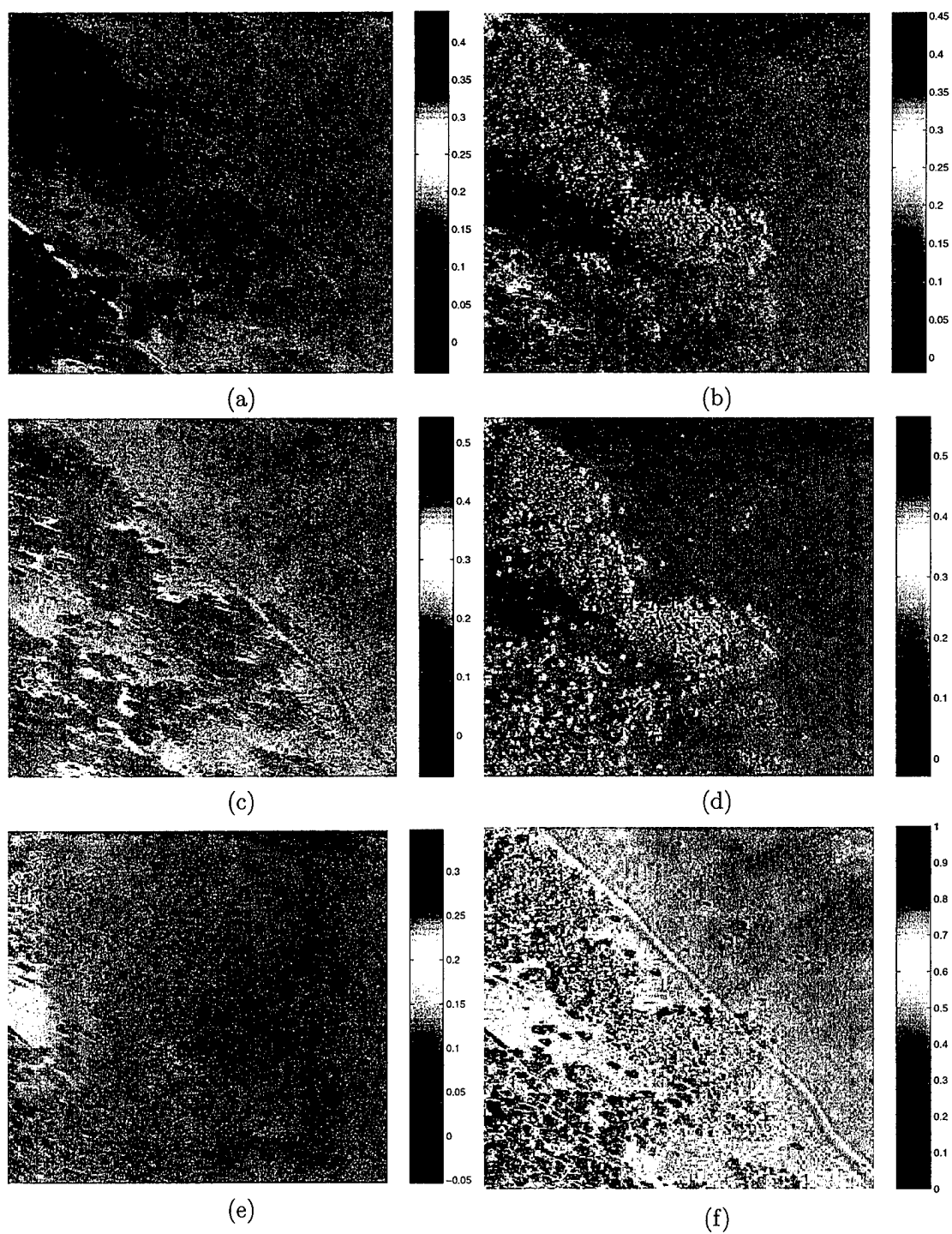


Figure 11. Abundance planes (a) Endmember 1; (b) Endmember 2; (c) Endmember 3; (d) Endmember 4; (e) Endmember 5; (f) Sum of abundances.

## 6. SUMMARY

We have introduced three algorithm taxonomies in Figures 3, 4, and 5 for hyperspectral unmixing whose organization begins with the features in Figure 2. The majority of algorithms are designed around a combination of three common assumptions: linear mixing, Gaussianity, and, the minimization of squared-error. A key discriminating feature of algorithms is their use of statistics, which has significant implications, especially in detection, when sub-pixel targets are comparatively rare and their discerning features may be lost in aggregate measures. Likewise, the use of parametric models to describe the random behavior of pixels is focused primarily on Gaussian models, but some efforts have been attempted to exploit higher-order statistics. While non-linear unmixing has been investigated by a number of researchers, it is still premature to develop a coherent taxonomy of algorithms in this field.

It is clear that methods for unmixing that incorporate a strong physical intuition have a better chance of successfully unmixing pixels in a hyperspectral image. We saw in our experiments that the geometric endmember determination algorithm was able to extract spectra that bore a strong resemblance, at least in shape, to distinct spectra in the scene. In contrast, the fuzzy clustering technique did not exhibit the ability to yield endmembers with a strong physical correlation to spectra in the scene. As a consequence, its abundance planes did not exhibit a strong ability to highlight distinct substances. More evaluation must be performed with other algorithms and with data that has been ground-truthed.

Unmixing is an important class of algorithm for hyperspectral processing. It is an important contributor to vital military applications such as terrain characterization and trafficability that lead to an intelligent preparation of a battlespace. The combination of physical modeling and mathematical techniques is capable of yielding information from scenes that was previously unattainable by other sensors. Without accurate ground truth, however, complete unmixing of scenes is difficult to corroborate. Nevertheless, unmixing holds the potential of revealing more from hyperspectral scenes than any other application.

## GLOSSARY

BDRF	Bi-direction reflectance function
BLUE	Best linear unbiased estimator
DPFT	Dark-point-fixed transform
DUSD	Deputy Under Secretary of Defense
EM	Expectation maximization
FPFT	Fixed-point-free transform
FVA	Filter vector algorithm
ICA	Independent component analysis
LMM	Linear mixing model
LSE	Least squares error
MAP	Maximum a posteriori
ML	Maximum likelihood
MNF	Maximum noise fraction
MSE	Mean squared error
MVT	Minimum volume transform
MVUE	Minimum variance unbiased estimator
NAPC	Noise adjusted principle components
ORASIS	Optical real-time adaptive spectral identification system
PCA	Principal component analysis
QICAR	Quasi-intrinsic Gaussian conditional autoregression
SMM	Stochastic mixing model
SNR	Signal-to-noise ratio
SSM	Salient selection module
SVD	Singular value decomposition
TBLS	Total block least squares

## REFERENCES

- [1] Jessica Bayliss, J. Anthony Gualtieri, and Robert F. Cromp. Analyzing hyperspectral data with independent component analysis. In *Proc. of the SPIE*, pages 133–143, 1998.
- [2] James C. Bezdek, Robert Ehrlich, and William Full. FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2-3):191–203, 1984.
- [3] Joseph W. Boardman. Geometrix mixture analysis of imaging spectrometry data. In *IGARSS*, pages 2369–2371, 1994.
- [4] Christoph C. Borel and Siegfried A.W. Gerstl. Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sensing of the Environment*, 47:403–416, 1994.
- [5] Jeffrey Bowles, John Antoniadis, Mark Baumbach, John Grossman, Daniel Haas, Peter Palmadesso, and John Stracka. Real time analysis of hyperspectral data sets using NRL's ORASIS algorithm. In *Proc. of the SPIE*, pages 38–45, May 1997.
- [6] Jeffrey Bowles, Peter Palmadesso, John Antoniadis, Mark Baumbach, and L.J. Rickard. Use of filter vectors in hyperspectral data analysis. In *Proc. of the SPIE*, pages 148–157, May 1995.
- [7] Samir Chettri and Nathan Netanyahu. Spectral unmixing of remotely sensed imagery using maximum entropy. In *Proc. of the SPIE*, volume 2962, pages 55–62, October 1997.
- [8] R.N. Clark and T.L. Roush. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophys. Res.*, 89(B7):6329–6340, 1984.
- [9] Maurice D. Craig. Minimum-volume transforms for remotely sensed data. *IEEE Trans. Geosci. and Rem. Sens.*, 32:99–109, 1994.
- [10] G.M. Foody and D.P. Cox. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership model and fuzzy membership functions. *Intl. Journal of Rem. Sens.*, 15:619–631, 1994.
- [11] Daniel R. Fuhrmann. A simplex shrink-wrap algorithm. In *Proc. of the SPIE*, volume 3701, pages 155–166, April 1999.
- [12] Andrew A. Green, Mark Berman, Paul Switzer, and Maurice D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. and Rem. Sens.*, 26:65–74, 1988.
- [13] Bruce Hapke. Bidirectional reflectance spectroscopy 1. theory. *Journal of Geophys. Res.*, 86(B4):3039–3054, 1981.
- [14] Bruce Hapke and Eddie Wells. Bidirectional reflectance spectroscopy 2. experiments and observations. *Journal of Geophys. Res.*, 86(B4):3055–3060, 1981.



- [15] Y. H. Hu, H. B. Lee, and F. L. Scarpace. Optimal linear spectral unmixing. *IEEE Trans. Geosci. and Rem. Sens.*, 37:639–644, 1999.
- [16] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall Signal Processing Series. Prentice Hall, 1993.
- [17] J.T. Kent and K.V. Mardia. Spatial classification using fuzzy membership models. *IEEE Pattern Anal. Machine Intell.*, 10:659–671, 1988.
- [18] Maurice S. Klein Gebbinck and Theo E. Schouten. Decomposition of mixed pixels. In *Proc. of the SPIE*, volume 2579, pages 104–111, April 1995.
- [19] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. Series in Automatic Computation. Prentice-Hall, 1974.
- [20] James B. Lee, Stephen Woodyatt, and Mark Berman. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Trans. Geosci. and Rem. Sens.*, 28:295–304, 1990.
- [21] S. Liangrocapart and M. Petrou. Mixed pixels classification. In *Proc. of the EUROPTO Conf. on Image and Signal Proc. for Remote Sensing*, pages 72–83, September 1998.
- [22] Dimitris G. Manolakis, Vinay K. Ingle, and Stephen M. Kogon. *Statistical and Adaptive Signal Processing*. Series in Electrical and Computer Engineering. McGraw-Hill, 2000.
- [23] T.K. Moon. The expectation-maximization algorithm. *IEEE Signal Processing Mag.*, 13(6), November 1996.
- [24] Terry Nichols, John Thomas, Woody Kober, and Vince Velten. A comparison of quantitative analysis techniques to discriminate intimately mixed bidirectional laboratory reflectance spectra. In *Proc. of the SPIE*, volume 3717, pages 2–12, April 1999.
- [25] Maria Petrou and Patricia G. Foschi. Confidence in linear spectral unmixing of single pixels. *IEEE Trans. Geosci. and Rem. Sens.*, 37:624–626, 1999.
- [26] Franco P. Preparata and Michael Ian Shamos. *Computational Geometry*. Springer-Verlag, 1985.
- [27] Jeff Settle and Norm Campbell. On the errors of two estimators of sub-pixel fractional cover when mixing is linear. *IEEE Trans. Geosci. and Rem. Sens.*, 36:163–170, 1998.
- [28] Donald L. Snyder, Joseph A. O'Sullivan, Daniel R. Fuhrmann, and William H. Smith. Estimation of overlapping spectral signatures from hyperspectral data. In *Proc. of the SPIE*, volume 3718, pages 470–479, April 1999.
- [29] S. S. Stan. Beyond the resolution limit: Using least squares for subpixel analysis in remote sensing. In *Proceedings of IEEE AFRICON '96*, volume 2, pages 597–602, September 1996.
- [30] Alan D. Stocker and Alan P. Schaum. Application of stochastic mixing models to hyperspectral detection problems. In *Proc. of the SPIE*, volume 3071, pages 47–60, September 1997.

- [31] Gilbert Strang. *Linear Algebra and Its Applications*. Harcourt Brace Jovanovich, 1988.
- [32] A. Tarantola and B. Valette. Generalized nonlinear inverse problems solved using the least squares criterion. *Review of Geophysics and Space Physics*, pages 20:219–232, 1982.
- [33] Stefanie Tompkins, John F. Mustard, Carle M. Pieters, and Donald W. Forsyth. Optimization of endmembers for spectral mixture analysis. *Remote Sensing of the Environment*, pages 472–489, 1997.
- [34] Joseph M. Triscari, A. Evan Iverson, Philip J. Sementilli, and Karen F. West. Projection-based methods for hyperspectral image fraction estimation. In *IEEE Southwest Symp. on Image Analysis and Interpretation*, pages 244–249, April 1998.
- [35] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory: Part I*. John Wiley and Sons, 1968.

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 15 January 2002		3. REPORT TYPE AND DATES COVERED Project Report	
4. TITLE AND SUBTITLE A Taxonomy of Spectral Unmixing Algorithms and Performance Comparisons				5. FUNDING NUMBERS  C — F19628-00-C-0002	
6. AUTHOR(S)  N. Keshava					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Lincoln Laboratory, MIT 244 Wood Street Lexington, MA 02420-9108				8. PERFORMING ORGANIZATION REPORT NUMBER  PR-HTAP-9	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) DUSD (S&T) Rosslyn Plaza North, Suite 9030 1777 N. Kent St. Rosslyn, VA 22209				10. SPONSORING/MONITORING AGENCY REPORT NUMBER  ESC-TR-2001-045	
11. SUPPLEMENTARY NOTES  None					
12a. DISTRIBUTION/AVAILABILITY STATEMENT  Approved for public release; distribution is unlimited.				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  In this report, algorithms for spectral unmixing are organized into taxonomies and their performance is then compared. Our motivation is to collectively organize and relate algorithms in order to assess the current state-of-the-art in the field and to facilitate objective comparisons between methods. The hyperspectral sensing community is populated by investigators with disparate scientific backgrounds, and efforts in spectral unmixing developed within disparate communities have inevitably led to duplication. This report is intended to remove ambiguity and redundancy by using a standard vocabulary, and clearly summarize what has and has not been done. As will be evident, the framework for the taxonomies derives its organization from the fundamental, philosophical assumptions imposed on the problem, rather than the common calculations they perform, or the similar outputs they might yield. The taxonomies are supplemented by a comparison of unmixing performance using techniques that typify the approaches of wide classes of algorithms.					
14. SUBJECT TERMS				15. NUMBER OF PAGES 58	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Same as Report	19. SECURITY CLASSIFICATION OF ABSTRACT Same as Report	20. LIMITATION OF ABSTRACT Same as Report		